

**KLASIFIKASI *SUPPORT VECTOR MACHINE* (SVM)  
TERHADAP SENTIMEN PUBLIK PADA KENAIKAN UANG  
KULIAH TUNGGAL (UKT)  
(STUDI KASUS: PLATFORM X)**

**SKRIPSI**

**OLEH:  
DEWI FARHANA MAULIYA  
NIM. 210601110099**



**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG  
2025**

**KLASIFIKASI *SUPPORT VECTOR MACHINE* (SVM)  
TERHADAP SENTIMEN PUBLIK PADA KENAIKAN UANG  
KULIAH TUNGGAL (UKT)  
(STUDI KASUS: PLATFORM X)**

**SKRIPSI**

**Diajukan Kepada  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
untuk Memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh  
Dewi Farhana Mauliya  
NIM. 210601110099**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG  
2025**

**KLASIFIKASI *SUPPORT VECTOR MACHINE* (SVM)  
TERHADAP SENTIMEN PUBLIK PADA KENAIKAN UANG  
KULIAH TUNGGAL (UKT)  
(STUDI KASUS: PLATFORM X)**

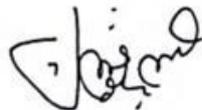
**SKRIPSI**

**Oleh  
Dewi Farhana Mauliya  
NIM. 210601110099**

Telah Disetujui Untuk Diuji

Malang, 19 Mei 2025

Dosen Pembimbing I



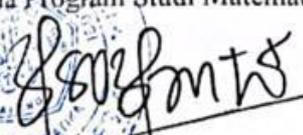
Ria Dhea Layla Nur Karisma, M.Si.  
NIPPPK. 19900709 202321 2 037

Dosen Pembimbing II



Juhari, M.Si.  
NIPPPK. 19840209 202321 1 010

Mengetahui,  
Ketua Program Studi Matematika



Dr. Jolly Susanti, M.Sc.  
NIP. 19741129 200012 2 005



**KLASIFIKASI *SUPPORT VECTOR MACHINE* (SVM)  
TERHADAP SENTIMEN PUBLIK PADA KENAIKAN UANG  
KULIAH TUNGGAL (UKT)  
(STUDI KASUS: PLATFORM X)**

**SKRIPSI**

Oleh  
**Dewi Farhana Mauliya**  
NIM. 210601110099

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)  
Tanggal 16 Juni 2025

Ketua Penguji : Dr. Fachrur Rozi, M.Si

Anggota Penguji 1 : Abdul Aziz, M.Si

Anggota Penguji 2 : Ria Dhea Layla Nur Karisma, M.Si

Anggota Penguji 3 : Juhari, M.Si



Mengetahui,  
Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc  
NIP. 19741129 200012 2 005

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Dewi Farhana Mauliya  
NIM : 210601110099  
Program Studi : Matematika  
Fakultas : Sains dan Teknologi  
Judul Skripsi : Klasifikasi *Support Vector Machine* (SVM) terhadap Sentimen Publik pada Kenaikan Uang Kuliah Tunggal (UKT) (Studi Kasus : Platform X)

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 16 Juni 2025

Yang membuat pernyataan,



Dewi Farhana Mauliya  
NIM. 210601110099

## **MOTO**

“Hari ini mungkin sulit, tapi hari esok adalah milik mereka yang tidak menyerah.”

## PERSEMBAHAN

Segala puji dan syukur penulis panjatkan kehadirat Allah SWT atas limpahan kasih sayang, kekuatan, dan kemudahan yang telah diberikan dalam menyelesaikan skripsi ini. Dengan segenap cinta dan ketulusan, skripsi ini penulis persembahkan kepada:

Keluarga tercinta, Bapak dan Ibu yang selalu menjadi cahaya dalam gelap, doa dalam diam, dan kekuatan dalam lelah. Terima kasih atas kasih sayang yang tak terhingga, pengorbanan yang tak ternilai, serta keyakinan yang tak pernah pudar terhadap langkah dan impian penulis. Untuk kakak-kakakku yang selalu memberikan dukungan, dan semangat kepada penulis.

Sahabat-sahabat terbaik. Terima kasih atas setiap pelukan hangat, semangat yang tak pernah habis, dan tawa yang menjadi obat kala penat melanda. Kalian adalah rumah kedua, tempat berbagi suka dan duka. Kebersamaan dan kehadiran kalian menjadi penopang dalam proses panjang ini, dan akan selalu menjadi bagian berharga dalam perjalanan hidup penulis.

Diri saya sendiri. Untuk semua air mata yang jatuh dalam diam, untuk setiap malam yang dilalui dengan rasa cemas dan lelah, dan untuk setiap pagi yang dimulai dengan tekad untuk terus melangkah. Terima kasih telah terus bertahan, terus belajar, dan tidak menyerah meskipun jalan terasa berbatu.

## KATA PENGANTAR

Puji dan syukur kehadiran Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi *Support Vector Machine* (SVM) terhadap Sentimen Publik pada Kenaikan Uang Kuliah Tunggal (UKT) (Studi Kasus: Platform X)”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Matematika pada Program Studi Matematika, Fakultas Sains dan teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Penulis menyadari bahwa penyusunan skripsi ini tidak akan terwujud tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim.
2. Prof. Dr. Sri Harini, M.Si, selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.
3. Dr. Elly Susanti, S.Pd., M.Sc, selaku ketua Program Studi Matematika, Universitas Islam Negeri Maulana Malik Ibrahim.
4. Ria Dhea Layla Nur Karisma, M.Si, selaku dosen pembimbing I yang dengan penuh kesabaran dan ketulusan membimbing, memberikan masukan, serta arahan sehingga skripsi ini dapat terselesaikan.
5. Juhari, M.Si, selaku dosen pembimbing II yang telah memberikan bimbingan, dukungan, dan motivasi yang sangat berarti bagi penulis dalam menyelesaikan skripsi ini.
6. Erna Herawati, M.Pd., selaku dosen wali yang telah memberikan bimbingan, dukungan, dan motivasi bagi penulis dalam menyelesaikan skripsi ini.
7. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim yang telah memberikan ilmu, wawasan, dan pengalaman berharga selama masa perkuliahan
8. Bapak Masturani, Ibu Rumlah, Kakak Prakas Ary Cahya, Kakak Mohammad Robby Alfirullah dan seluruh keluarga besar yang selalu memberikan doa, dukungan fisik, mental serta finansial, dan semangat kepada penulis hingga

selesainya skripsi ini. Luka sekaligus obat yang diberikan sangat berarti pada penulis untuk dapat menyelesaikan skripsi ini.

9. Kepada Ahmad Lamim Putra yang berkontribusi banyak dalam penulisan skripsi ini, baik tenaga, waktu maupun materi kepada penulis sekaligus keluarga yang selalu memberi semangat dan perhatian kepada penulis.
10. Kepada sahabat-sahabat penulis, yaitu Karin, Reja, Elok, Ailsa, Faradiva, Arsita, serta teman-teman lainnya yang tidak dapat disebutkan satu per satu, terima kasih atas dukungan, kebersamaan, dan semangat yang diberikan.
11. Kepada teman-teman semasa SMA dan SMP yang kini tidak dapat lagi bersama, terima kasih atas doa, kehangatan serta impian-impian yang dahulu pernah diimpikan bersama.
12. Seluruh mahasiswa angkatan 2021.

Malang, 13 Juni 2025

Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PENGAJUAN .....	ii
HALAMAN PERSETUJUAN .....	iii
HALAMAN PENGESAHAN .....	iv
PERNYATAAN KEASLIAN TULISAN.....	v
MOTO .....	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL .....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR SIMBOL .....	xiv
DAFTAR LAMPIRAN .....	xv
ABSTRAK .....	xvi
ABSTRACT .....	xvii
مستخلص البحث.....	xviii
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	6
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	6
1.5 Batasan Masalah.....	7
<b>BAB II KAJIAN TEORI .....</b>	<b>8</b>
2.1 Teori Pendukung .....	8
2.1.1 Analisis Sentimen.....	8
2.1.2 Klasifikasi Teks .....	8
2.1.3 <i>Text Preprocessing</i> .....	11
2.1.4 Visualisasi Data .....	13
2.1.5 <i>Imbalanced Data</i> .....	13
2.1.6 Ekstraksi Fitur .....	14
2.1.7 <i>Support Vector Machine (SVM)</i> .....	16
2.1.8 <i>Grid Search</i> dan <i>Cross Validation</i> .....	23
2.1.9 Evaluasi Model.....	25
2.1.10 Uang Kuliah Tunggal (UKT) .....	28
2.1.11 Platform X .....	29
2.2 Kajian Integrasi Topik dengan Al-Quran/Hadits .....	30
2.3 Kajian Topik dengan Teori Pendukung .....	32
<b>BAB III METODE PENELITIAN .....</b>	<b>35</b>
3.1 Jenis Penelitian.....	35
3.2 Data dan Sumber Data.....	35
3.3 Teknik Pengumpulan Data.....	35
3.4 Teknik Analisis Data.....	36
3.2 Diagram Alir Penelitian .....	40

<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>41</b>
4.1 Pengumpulan Data .....	41
4.2 Pelabelan Data.....	42
4.3 <i>Text Preprocessing</i> .....	43
4.4 Visualisasi Data.....	47
4.5 Pembagian Data.....	48
4.6 Klasifikasi SVM.....	49
4.6.1 Klasifikasi SVM Manual terhadap Sampel Data.....	49
4.6.2 Klasifikasi SVM terhadap Seluruh Data .....	58
4.7 Analisis Sentimen terhadap Kebijakan Kenaikan Uang Kuliah Tunggal dalam Pandangan Islam .....	68
<b>BAB V PENUTUP.....</b>	<b>70</b>
5.1 Kesimpulan.....	70
5.2 Saran.....	71
<b>DAFTAR PUSTAKA .....</b>	<b>72</b>
<b>LAMPIRAN.....</b>	<b>75</b>
<b>RIWAYAT HIDUP .....</b>	<b>88</b>

## DAFTAR TABEL

Tabel 2.1	Contoh Kalimat dalam Dokumen .....	15
Tabel 2.2	Contoh Perhitungan <i>Term Frequency</i> (TF).....	16
Tabel 2.3	Contoh Perhitungan <i>Inverse Document Frequency</i> (IDF) .....	16
Tabel 2.4	Contoh Perhitungan TF-IDF .....	16
Tabel 2.5	<i>Confusion Matrix</i> .....	25
Tabel 2.6	Nilai AUC dan Interpretasi .....	28
Tabel 3.1	Kriteria Pelabelan Data .....	36
Tabel 4.1	Sampel Data <i>Tweet</i> .....	43
Tabel 4.2	Hasil Proses <i>Case Folding</i> Data <i>Tweet</i> .....	44
Tabel 4.3	Hasil Proses <i>Cleaning</i> Data <i>Tweet</i> .....	44
Tabel 4.4	Hasil Proses <i>Normalisasi</i> Data <i>Tweet</i> .....	45
Tabel 4.5	Hasil Proses <i>Tokenizing</i> Data <i>Tweet</i> .....	45
Tabel 4.6	Hasil Proses <i>Stopwords Removal</i> Data <i>Tweet</i> .....	46
Tabel 4.7	Hasil Proses <i>Stemminng</i> Data <i>Tweet</i> .....	46
Tabel 4.8	Akurasi Perbandingan Data <i>Training</i> dan Data <i>Testing</i> .....	48
Tabel 4.9	Pembagian Data <i>Training</i> dan <i>Testing</i> .....	48
Tabel 4.10	Sampel Data <i>Tweet</i> .....	49
Tabel 4.11	Hasil Perhitungan Frekuensi Kemunculan Term, TF dan IDF(t) ..	50
Tabel 4.12	Hasil Perhitungan TF-IDF .....	51
Tabel 4.13	Data <i>Training</i> .....	53
Tabel 4.14	Data <i>Testing</i> .....	57
Tabel 4.15	Nilai Akurasi Parameter Model SVM <i>Kernel Linear</i> .....	60
Tabel 4.16	Nilai Akurasi Parameter Model SVM <i>Kernel RBF</i> .....	60
Tabel 4.17	Nilai Akurasi Parameter Model SVM <i>Kernel Sigmoid</i> .....	61
Tabel 4.18	Nilai Akurasi Parameter Model SVM <i>Kernel Polinomial</i> .....	61
Tabel 4.19	Perbandingan Performa Setiap <i>Kernel</i> .....	62
Tabel 4.20	<i>Confusion Matrix</i> .....	62
Tabel 4.21	Nilai FPR dan TPR .....	65

## DAFTAR GAMBAR

Gambar 1.1 Jumlah Mahasiswa di Indonesia.....	3
Gambar 2.1 Contoh <i>Word Cloud</i> .....	13
Gambar 2.2 Ilustrasi SVM .....	17
Gambar 2.3 Proses Transformasi Data dalam Ruang Dimensi Tinggi .....	21
Gambar 3.1 <i>Flowchart</i> Tahap Analisis Data.....	40
Gambar 4.1 Perbandingan Jumlah <i>Tweet</i> Berdasarkan Kategori Sentimen.....	42
Gambar 4.2 Visualisasi Data <i>Tweet</i> Terkait Kebijakan Kenaikan UKT.....	47
Gambar 4.3 Perbandingan Pembagian Data Sebelum dan Sesudah SMOTE..	59
Gambar 4.4 Kurva ROC dan Nilai AUC Model SVM .....	65

## DAFTAR SIMBOL

$\alpha$	: <i>Lagrange Multiplier</i>
$\mathbf{w}$	: Vektor Pembobot
$b$	: Bias
$K(x_i, x_j)$	: Fungsi <i>Kernel</i>
$f(t, D)$	: Frekuensi kata $t$ dalam dokumen $D$
$df(t)$	: Jumlah dokumen yang memuat kata $t$ .
$tf$	: Frekuensi kemunculan kata dalam dokumen
$idf$	: <i>Inverse</i> frekuensi dalam dokumen
$\xi_i$	: <i>Slack variables</i>
$y_i \in \{1, -1\}$	: Label kelas
$\mathbf{x}_i$	: Data ke - $i$
$\mathbf{x}_j$	: Data ke - $j$
$m$	: Jumlah fitur dalam setiap data
$i, j$	: Urutan data
$l$	: Jumlah data

## DAFTAR LAMPIRAN

Lampiran 1. <i>Script Crawling Data X</i> .....	75
Lampiran 2. <i>Script Preprocessing text</i> .....	75
Lampiran 3. <i>Script Visualisasi Data</i> .....	78
Lampiran 4. <i>Script Klasifikasi SVM</i> .....	79
Lampiran 5. <i>Kamus Slang Word</i> .....	81
Lampiran 6. <i>Kamus Stopword Text</i> .....	81
Lampiran 7. <i>Pelabelan Data</i> .....	82
Lampiran 8. <i>Data Preprocessing Text</i> .....	83

## ABSTRAK

Mauliya, Dewi Farhana. 2025. **Klasifikasi Support Vector Machine (SVM) terhadap Sentimen Publik pada Kenaikan Uang Kuliah Tunggal (UKT) (Studi Kasus: Platform X)**. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Ria Dhea Layla Nur Karisma, M.Si. (II) Juhari, M.Si.

**Kata kunci:** Sentimen Publik, Uang Kuliah Tunggal, Klasifikasi, *Support Vector Machine*, Akurasi

Kebijakan kenaikan Uang Kuliah Tunggal (UKT) di perguruan tinggi negeri memicu berbagai reaksi dari masyarakat, khususnya di media sosial. Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap kebijakan tersebut dengan menggunakan metode *Support Vector Machine* (SVM). Data yang digunakan diperoleh dari Platform X dan terdiri atas *tweet* yang mengandung opini publik terkait kenaikan UKT. Tahapan penelitian meliputi pengumpulan data, *preprocessing*, pelabelan sentimen secara manual, ekstraksi fitur menggunakan metode TF-IDF, pelatihan model SVM, serta evaluasi performa model. Berdasarkan hasil klasifikasi terhadap 197 data *testing*, model SVM berhasil mengklasifikasikan 182 data sebagai sentimen negatif dan 2 data sebagai sentimen positif secara tepat. Namun demikian, terdapat 13 data yang diklasifikasikan secara keliru, yaitu 2 data negatif yang terklasifikasi sebagai positif dan 11 data positif yang terklasifikasi sebagai negatif. Evaluasi performa model menunjukkan tingkat akurasi sebesar 93% dan nilai ROC AUC sebesar 74%, yang mengindikasikan bahwa model memiliki kinerja yang baik secara umum.

## ABSTRACT

Mauliya, Dewi Farhana. 2025. **Support Vector Machine (SVM) Classification of Public Sentiment on Single Tuition Fee (UKT) Increases (Case Study: Platform X)**. Thesis. Mathematics Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Ria Dhea Layla Nur Karisma, M.Si. (II) Juhari, M.Si.

**Keywords:** Public Sentiment, Single Tuition Fee, Classification, Support Vector Machine, Accuracy

The policy of increasing the Single Tuition Fee (UKT) in public universities has triggered various reactions from the public, especially on social media. This research aims to classify public sentiment towards the policy using the Support Vector Machine (SVM) method. The data used is obtained from Platform X and consists of uploads containing public opinion related to the UKT increase. The research stages include data collection, preprocessing, manual sentiment labeling, feature extraction using TF-IDF method, SVM model training, and model performance evaluation. Based on the classification results of 197 test data, the SVM model successfully classified 182 data as negative sentiment and two data as positive sentiment correctly. However, there are 13 data classified incorrectly, namely two negative data classified as positive and 11 positive data classified as negative. The performance evaluation of the model shows an accuracy rate of 93% and a ROC AUC value of 74%, which indicates that the model has good performance in general.

## مستخلص البحث

ماوليا، دوي فرحنا. ٢٠٢٥. تصنيف آلة الدعم المتجه (SVM) تجاه الرأي العام بشأن زيادة الرسوم الدراسية الموحدة (UKT) (دراسة حالة: منصة X). البحث العلمي. قسم الرياضيات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرفة: (١) ربا ديا ليلي نور كاريسما، الماجستير في العلوم. (٢) جوهاري، الماجستير في العلوم.

الكلمات الأساسية: الرأي العام، الرسوم الدراسية الموحدة، التصنيف، آلة الدعم المتجهي، الدقة

سياسة زيادة الرسوم الدراسية الموحدة (UKT) في الجامعات الحكومية أثارت ردود فعل متنوعة من المجتمع، خاصة على وسائل التواصل الاجتماعي. تهدف هذه الدراسة إلى تصنيف الرأي العام تجاه هذه السياسة باستخدام طريقة آلة الدعم المتجه (SVM). تم الحصول على البيانات المستخدمة من منصة X وتشمل تغريدات تحتوي على آراء عامة بشأن زيادة الرسوم الدراسية الموحدة. تشمل مراحل البحث جمع البيانات، والمعالجة المسبقة، وتصنيف المشاعر يدويًا، واستخراج الميزات باستخدام طريقة TF-IDF، وتدريب نموذج SVM، وتقييم أداء النموذج. استنادًا إلى نتائج تصنيف ١٩٧ بيانات اختبار، نجح نموذج SVM في تصنيف ١٨٢ بيانات على أنها مشاعر سلبية و ٢ بيانات على أنها مشاعر إيجابية بشكل دقيق. ومع ذلك، هناك ١٣ بيانات تم تصنيفها بشكل خاطئ، وهي ٢ بيانات سلبية تم تصنيفها على أنها إيجابية و ١١ بيانات إيجابية تم تصنيفها على أنها سلبية. أظهر تقييم أداء النموذج دقة بنسبة ٩٣٪ وقيمة ROC AUC بنسبة ٧٤٪، مما يشير إلى أن النموذج يتمتع بأداء جيد بشكل عام.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Analisis sentimen merupakan proses yang bertujuan untuk menganalisis opini, sikap, atau emosi yang terkandung dalam teks. Tujuan utama dari analisis ini adalah untuk mengklasifikasikan sentimen sebagai positif atau negatif, yang berpotensi memberikan wawasan mendalam mengenai opini publik terkait suatu isu (Herlinawati dkk., 2020). Sentimen merupakan pandangan atau perasaan yang muncul sebagai hasil dari persepsi terhadap suatu fenomena. Seiring dengan berkembangnya teknologi, analisis sentimen menjadi alat yang sangat penting untuk memahami opini publik, terutama dalam data teks yang tidak terstruktur. Opini publik yang beragam dan besar dapat menyulitkan serta memerlukan waktu yang lama untuk dianalisis secara langsung. Oleh karena itu, diperlukan metode yang mampu menganalisis dan mengklasifikasikan opini tersebut ke dalam kelas-kelas tertentu.

Salah satu metode yang terbukti efektif dalam menganalisis data sentimen yang besar dan beragam adalah dengan memanfaatkan *Support Vector Machine* (SVM). SVM merupakan algoritma klasifikasi dalam *supervised learning* yang digunakan untuk memisahkan dua kelas dengan mencari *hyperplane* terbaik dalam ruang fitur data. Keunggulan SVM terletak pada kemampuannya dalam memaksimalkan margin antar kelas, sehingga menghasilkan tingkat akurasi yang tinggi dalam klasifikasi sentimen, baik positif maupun negatif (Nurhafida & Sembiring, 2022).

Proses klasifikasi SVM didasarkan pada prinsip bahwa semakin lebar margin, maka semakin kecil kesalahan generalisasi dari pemisah tersebut. SVM secara matematis mencari *hyperplane* terbaik dari berbagai kemungkinan untuk memaksimalkan margin, sehingga menghasilkan model dengan performa yang baik pada data baru (Rangga dkk., 2019). SVM digunakan untuk menyelesaikan permasalahan klasifikasi, baik yang bersifat *linear* maupun *non-linear*. SVM juga menyediakan berbagai jenis *kernel* yang dapat disesuaikan untuk memenuhi kebutuhan proses klasifikasi tertentu (Yu dkk., 2010). Keunggulan SVM tersebut sangat berguna dalam menganalisis sentimen masyarakat terhadap berbagai isu, termasuk kebijakan pendidikan.

Salah satu kebijakan akademik yang menjadi perhatian publik adalah kebijakan Uang Kuliah Tunggal (UKT). UKT, sebagai bagian dari Biaya Kuliah Tunggal (BKT), adalah biaya yang harus dibayar oleh mahasiswa berdasarkan kemampuan ekonomi orang tua atau wali. Berdasarkan Permendikbud Nomor 2 tahun 2024, Perguruan Tinggi Negeri (PTN) memiliki kewenangan untuk menetapkan UKT sesuai BKT di setiap program studi. Namun, dalam implementasinya, kebijakan ini sering kali menimbulkan beban finansial bagi sebagian keluarga, khususnya dengan adanya kebijakan kenaikan UKT yang tidak sesuai dengan finansial keluarga. Kenaikan UKT ini memicu respons publik yang beragam, mulai dari kritik hingga dukungan terhadap kebijakan tersebut.

Tren kenaikan jumlah mahasiswa di Indonesia semakin memperkuat pentingnya kebijakan UKT yang transparan dan adil. Data dari Badan Pusat Statistika (BPS) pada Gambar 1.1, menunjukkan peningkatan jumlah mahasiswa dari 6,9 juta pada tahun 2018 menjadi 9,2 juta pada tahun 2022 (BPS, 2023).

Peningkatan jumlah mahasiswa mencerminkan minat yang tinggi terhadap pendidikan tinggi di Indonesia. Sehingga, kebijakan UKT yang transparan dan adil sangat penting agar semua lapisan masyarakat dapat mengakses pendidikan.



**Gambar 1.1** Jumlah Mahasiswa di Indonesia (BPS, 2023)

Respons publik terhadap kenaikan UKT semakin meningkat di berbagai platform media sosial, termasuk Platform X. Platform X merupakan media sosial yang sebelumnya dikenal dengan nama Twitter. Platform ini menjadi tempat utama bagi publik untuk mengekspresikan opini mereka terkait kebijakan pendidikan, termasuk kenaikan UKT. Pengguna Platform X mengkritik kebijakan tersebut, dengan alasan bahwa biaya kuliah yang semakin tinggi menambah beban keluarga, terutama dalam kondisi ekonomi yang tidak sesuai dengan golongan UKT.

Dalam konteks pengambilan kebijakan yang berdampak pada banyak orang, penting untuk menerapkan prinsip keadilan dan transparansi. Tercermin dalam firman Allah SWT dalam QS. Al-Baqarah: 42, yaitu:

وَلَا تَلْبِسُوا الْحَقَّ بِالْبَاطِلِ وَتَكْتُمُوا الْحَقَّ وَأَنْتُمْ تَعْلَمُونَ

(Kementerian Agama RI, 2022)

Artinya: *“Dan janganlah kamu campur adukkan yang hak dengan yang bathil dan janganlah kamu sembunyikan yang hak itu, sedang kamu mengetahui.”*

Menurut Ibnu Katsir, ayat ini menegaskan pentingnya keterbukaan, kejujuran, dan transparansi dalam setiap tindakan dan kebijakan yang diambil (Ghoffar dkk., 2004). Kebijakan seperti UKT, yang berdampak besar pada masyarakat, harus ditetapkan secara adil dan terbuka. Kebijakan yang tidak transparan berpotensi menimbulkan ketidakpercayaan dan ketidakpuasan di kalangan masyarakat. Oleh karena itu, sangat penting bagi pihak yang berwenang untuk menghindari pencampuran antara yang benar dan yang salah dalam menetapkan kebijakan, serta memastikan bahwa semua keputusan didasarkan pada prinsip keadilan yang sesuai dengan syariat.

Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia (2021-2024), Nadiem Makarim, menunda kenaikan UKT pada tahun 2024 sebagai respons atas kekhawatiran masyarakat (Mawardi, 2024). Kebijakan ini diambil setelah melihat banyaknya kritik yang berkembang di platform media sosial, termasuk Platform X, yang menjadi kanal utama bagi mahasiswa dan masyarakat umum untuk menyuarakan pendapat mereka. Dengan ribuan *tweet* dan komentar yang tersebar di platform tersebut, diperlukan metode yang efisien untuk menganalisis sentimen publik secara tepat.

Kebijakan pemerintah sering memicu berbagai respons publik. Oleh karena itu, analisis sentimen diperlukan untuk memahami persepsi masyarakat mengenai kebijakan kenaikan UKT. Berdasarkan penelitian sebelumnya, SVM telah terbukti memiliki performa yang baik dalam klasifikasi sentimen publik dari platform X, sehingga penelitian ini menggunakan SVM untuk menganalisis sentimen

masyarakat terhadap kebijakan kenaikan UKT di platform tersebut. Dengan memahami dan mengklasifikasikan sentimen pengguna, penelitian ini diharapkan dapat memberikan gambaran yang jelas mengenai persepsi publik. Kemampuan SVM dalam menganalisis data besar dan beragam memungkinkan klasifikasi sentimen publik menjadi positif atau negatif secara akurat.

Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap kebijakan kenaikan UKT menggunakan metode SVM. Penelitian sebelumnya oleh Ariyanto & Chamidah (2021) menganalisis sentimen terkait kebijakan penerimaan melalui sistem zonasi dengan SVM dan Naive Bayes. Dalam penelitian tersebut, SVM menunjukkan akurasi yang lebih tinggi, yaitu 92,93% dibandingkan Naive Bayes yang memiliki akurasi 79,86%, sehingga SVM lebih efektif dalam analisis sentimen kebijakan zonasi. Penelitian lainnya dilakukan oleh Baita dkk. (2021) dalam analisis sentimen mengenai vaksin Sinovac menggunakan SVM dan KNN. Hasil penelitian tersebut menunjukkan bahwa SVM memiliki performa yang lebih baik jika dibandingkan dengan algoritma KNN. SVM memiliki tingkat akurasi sebesar 70%, sedangkan KNN sebesar 56%. Selain itu, penelitian lain dilakukan oleh Ramadhan & Ramadhan (2022), menunjukkan bahwa SVM unggul dengan akurasi 79% dibandingkan metode Regresi Logistik dalam menganalisis sentimen terhadap *review* film pada komentar pengguna *website* IMDB. Penelitian serupa dilakukan oleh Wahyuni dkk. (2024) terkait analisis sentimen di youtube terhadap kenaikan UKT menggunakan Metode SVM. Hasil penelitian tersebut menunjukkan bahwa SVM mencapai akurasi 88% dengan data yang diperoleh yaitu 121

komentar. Sehingga, peneliti memilih platform X untuk digunakan sebagai platform pengambilan data agar data yang diperoleh besar dan cukup beragam.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan, dapat dirumuskan masalah pada penelitian ini, yaitu:

1. Bagaimana klasifikasi sentimen publik terhadap kebijakan kenaikan UKT di Platform X menggunakan SVM?
2. Bagaimana tingkat akurasi dari klasifikasi sentimen publik terhadap kebijakan kenaikan UKT di Platform X menggunakan SVM?

## **1.3 Tujuan Penelitian**

Berdasarkan uraian rumusan masalah yang dipaparkan, dapat diperoleh tujuan penelitian, yaitu:

1. Mengetahui klasifikasi sentimen publik terhadap kebijakan kenaikan UKT di Platform X menggunakan SVM.
2. Mengetahui tingkat akurasi dari klasifikasi sentimen publik terhadap kebijakan kenaikan UKT di Platform X menggunakan SVM.

## **1.4 Manfaat Penelitian**

1. Bagi Mahasiswa: mahasiswa memperoleh pengalaman dalam menerapkan metode analisis sentimen dan algoritma klasifikasi seperti SVM, yang akan meningkatkan pemahaman tentang *machine learning* dan analisis data.

2. Bagi Pemerintah: pemerintah dapat menggunakan hasil penelitian sebagai dasar informasi untuk mengevaluasi kebijakan kenaikan UKT dan mempertimbangkan penyesuaian kebijakan pendidikan yang lebih sensitif terhadap kondisi dan aspirasi masyarakat.
3. Bagi Program Studi Matematika: penelitian yang dilakukan oleh mahasiswa dapat menambah studi literatur matematika di bidang statistika khususnya dalam analisis data dan *machine learning*.

### **1.5 Batasan Masalah**

Batasan masalah dari penelitian ini, yaitu:

1. Data yang digunakan dalam penelitian ini diperoleh dari Platform X dalam bentuk 1.000 *tweet* yang membahas kenaikan UKT.
2. Hanya *tweet* berbahasa Indonesia yang digunakan dalam penelitian ini, sedangkan *tweet* dalam bahasa lain tidak termasuk dalam analisis.
3. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*, ROC AUC tanpa mempertimbangkan metrik evaluasi lainnya.

## **BAB II**

### **KAJIAN TEORI**

#### **2.1 Teori Pendukung**

##### **2.1.1 Analisis Sentimen**

Analisis sentimen merupakan cabang ilmu yang berhubungan dengan *Natural Language Processing* (NLP). Tujuan dari analisis sentimen adalah untuk memahami, mengelola, menganalisis, dan mengekstrak data teks seperti opini, evaluasi, sikap, emosi terhadap topik tertentu. Manusia memiliki enam emosi dasar yaitu emosi takut (*fear*), marah (*anger*), jijik (*disgust*), sedih (*sadness*), terkejut (*surprise*), dan senang (*joy*) (Gudkk., 2019). Proses ini melibatkan pengolahan kalimat atau dokumen menjadi kata-kata yang memiliki makna, yang kemudian diklasifikasikan ke dalam sentimen positif dan sentimen negatif (Ekdkk., 2019).

Analisis sentimen dapat digunakan untuk mengamati tanggapan masyarakat terhadap berbagai isu, sehingga berguna dalam mengetahui respons politik (Yudkk., 2010). Reaksi, komentar, dan pendapat yang dianalisis melalui sentimen dapat mencerminkan seberapa efisien dan efektif kebijakan yang diberlakukan. Analisis sentimen memberikan wawasan berharga bagi pembuat kebijakan untuk menilai dan menyesuaikan strategi mereka.

##### **2.1.2 Klasifikasi Teks**

Klasifikasi teks adalah proses pengelompokan data teks ke dalam kategori atau kelas tertentu berdasarkan pola yang dipelajari dari data sebelumnya (Sarkar,

2016). Klasifikasi teks secara manual dapat dilakukan oleh manusia dengan cara membaca dan menganalisis setiap dokumen. Namun, apabila jumlah data yang dianalisis sangat besar, teknik *machine learning* dapat digunakan untuk membangun sistem klasifikasi teks secara otomatis.

*Machine learning* memiliki dua jenis teknik utama, yaitu *supervised learning* dan *unsupervised learning*. *Unsupervised learning* merujuk pada algoritma yang tidak memerlukan data *training* berlabel untuk membangun model. Sebaliknya, *supervised learning* melibatkan algoritma yang dilatih menggunakan data berlabel. Proses klasifikasi secara matematis dapat dituliskan sebagai berikut (Sarkar, 2016):

$$TN = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$$

di mana  $x_1, x_2, \dots, x_l$  adalah data teks, dan kelas yang sesuai adalah  $y_1, y_2, \dots, y_l$ . Misalkan algoritma  $Z$  dilatih menggunakan data *training*  $TN$ , sehingga menghasilkan model klasifikasi atau *classifier*  $\gamma$ , dan proses ini dikenal sebagai proses pelatihan.

Setelah model terbentuk, model tersebut dapat digunakan untuk memprediksi kelas dari dokumen baru yang belum pernah dilihat sebelumnya, misalnya dokumen  $TS$ , dan memprediksi kelas  $y_{TS}$ . Proses ini dapat direpresentasikan sebagai berikut pada persamaan (2.1) (Sarkar, 2016):

$$\gamma : TS \rightarrow y_{TS} \tag{2.1}$$

Dengan demikian, terdapat beberapa tahapan utama dalam algoritma klasifikasi, yaitu (Sarkar, 2016):

### 1. Pelatihan ( *Training* )

Sebelum proses pelatihan, dilakukan ekstraksi fitur untuk mendapatkan fitur-fitur bermakna dari data mentah. Fitur - fitur ini kemudian dimasukkan ke dalam algoritma yang dipilih, yang bertujuan untuk mengidentifikasi dan mempelajari pola dari fitur-fitur tersebut beserta hasil yang terkait. Hasil dari proses ini adalah model klasifikasi, yang diharapkan mampu memprediksi kelas untuk data baru di masa mendatang.

### 2 Evaluasi ( *Evaluation* )

Evaluasi dilakukan untuk menguji kinerja prediksi model serta mengukur seberapa baik model tersebut mempelajari pola dari data *training*. Pada tahap ini, digunakan data validasi untuk menguji prediksi model dan membandingkannya dengan label kelas aktual. Selain itu, metode *cross-validation* dapat diterapkan, di mana data dibagi menjadi beberapa bagian, dengan sebagian digunakan untuk pelatihan dan sebagian lainnya untuk validasi.

### 3 Penyetelan ( *Tuning* )

Tahap penyempurnaan model, yang juga dikenal sebagai penyetelan *hiperparameter*, berfokus pada optimasi model untuk memaksimalkan akurasi prediksi dan mengurangi kesalahan. *Hiperparameter* seperti parameter regularisasi atau tipe *kernel* pada SVM dapat disesuaikan untuk meningkatkan kinerja model dalam klasifikasi teks.

### 2.1.3 Text Preprocessing

*Text preprocessing* adalah tahap awal dalam persiapan data untuk pengolahan data klasifikasi. Pada klasifikasi teks, tahap *preprocessing* bertujuan untuk meningkatkan akurasi klasifikasi data. *Preprocessing* adalah proses pengolahan teks yang digunakan untuk menyeragamkan bentuk teks tidak terstruktur menjadi terstruktur sesuai dengan kebutuhan pengolahan data lebih lanjut. Proses *Preprocessing* cukup kompleks karena Bahasa Indonesia memiliki aturan penulisan dan pembentukan kata yang bervariasi. Aturan ini mencakup berbagai imbuhan yang mengubah makna kata dasar, seperti prefiks (awalan), sufiks (akhiran), konfiks atau simulfiks (awalan dan akhiran), dan infiks (sisipan) (Tando & Irawan, 2023). Tahapan dalam *preprocessing* adalah sebagai berikut:

1. *Case Folding*

Tahap *case folding* adalah proses mengubah semua huruf pada kalimat menjadi huruf kecil. Dalam melakukan analisis sentimen, setiap huruf dalam data sangat berpengaruh dan menentukan hasil analisis, sehingga perlu dilakukan perubahan menjadi huruf kecil agar proses analisis selanjutnya maksimal.

2. *Cleaning*

Tahap *cleaning* adalah proses penting dalam pengolahan data teks yang bertujuan untuk membersihkan data dari elemen-elemen yang tidak diperlukan. Proses penghapusan karakter dalam teks yang bukan bagian dari alfabet dan tidak memiliki makna tertentu, seperti angka, *hashtag*, *emoji*, *link*, tanda baca, dan lain sebagainya. Proses ini penting untuk mengurangi “*noise*”

dalam data, sehingga hanya menyisakan teks yang bersih dan relevan untuk tahap analisis selanjutnya.

### 3. Normalisasi

Proses normalisasi yaitu proses perubahan kata slang atau tidak baku diubah menjadi bentuk baku atau normal. Proses ini dilakukan dengan menggunakan kamus kata normalisasi yang berisi pasangan kata tidak baku dan baku. Misalnya, kata “se7” dinormalisasi menjadi kata “setuju”; kata “adlah” menjadi “adalah”, dan lain sebagainya.

### 4. *Tokenizing*

Tahap *tokenizing* adalah proses mengubah kalimat menjadi kumpulan kata atau bagian yang lebih kecil sebelum dianalisis lebih lanjut. Sebagai contoh, kalimat “kenaikan biaya ukt” akan dipisahkan menjadi kata-kata “kenaikan”, “biaya”, dan “ukt”. Proses *tokenizing* dilakukan dengan menggunakan *library nltk* yang tersedia di *Python*.

### 5. *Stopwords Removal*

Tahap *stopwords* adalah proses menghapus kata-kata umum yang tidak berkontribusi secara signifikan dalam analisis, sehingga mengurangi jumlah kata yang disimpan oleh sistem. Seperti kata “saya”, “dia”, “bawah”, “dengan”, “nya”, “itu”, “ini”, dan kata lainnya.

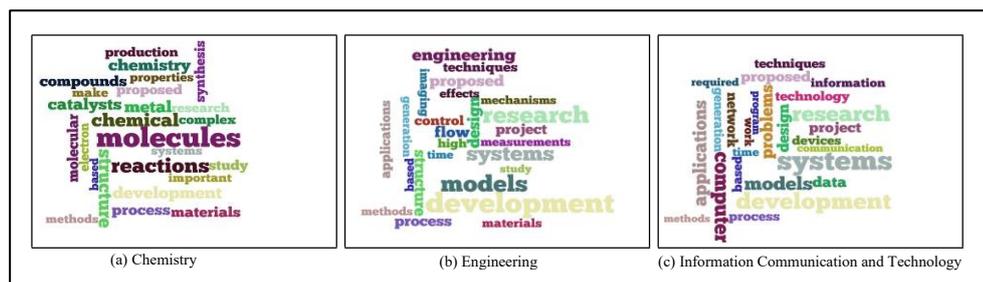
### 6. *Stemming*

Tahap *stemming* adalah proses mencari *root* (kata dasar) dari setiap kata hasil penghapusan *stopword*. Proses *stemming* menggunakan *library Sastrawi* dalam *Python* dan bertujuan untuk mengubah kata ke bentuk dasarnya sesuai dengan aturan dalam Kamus Besar Bahasa Indonesia (KBBI). Misalnya, kata

“menunggu” diubah menjadi kata dasar “tunggu”, “membantu” menjadi “bantu”, dan lain sebagainya.

#### 2.1.4 Visualisasi Data

Tahap visualisasi menggunakan *word cloud* adalah metode untuk menggambarkan secara grafis kata-kata yang paling sering muncul dalam sebuah dokumen teks. Proses ini dilakukan dengan menempatkan kata-kata dalam sebuah gambar atau plot dua dimensi, di mana ukuran relatif dari setiap kata menunjukkan seberapa sering kata tersebut muncul dalam teks tersebut. Semakin besar ukuran huruf kata dalam *word cloud* menunjukkan frekuensi kemunculan yang lebih tinggi. Teknik ini membantu dengan cepat mengidentifikasi kata-kata kunci atau tema utama yang dominan dalam dokumen tersebut, serta memberikan gambaran visual yang intuitif tentang konten teks tanpa perlu membaca secara detail setiap kata. Berikut merupakan contoh visualisasi dokumen teks menggunakan *word cloud* (Castella & Sutton, 2014)



Gambar 2.1 Contoh *Word Cloud*

#### 2.1.5 Imbalanced Data

Permasalahan yang sering muncul dalam analisis sentimen, yaitu ketidakseimbangan (*imbalanced data sets*) terkait jumlah tiap kelas, seperti

kecenderungan terhadap kelas positif atau sebaliknya. Salah satu metode yang populer dalam mengatasi ketidakseimbangan ini adalah metode *Synthetic Minority Over-sampling Technique* (SMOTE). Metode ini bekerja dengan membuat replikasi data dari kelas minoritas, yang disebut sebagai data sintesis (*synthetic data*) (Chawla dkk., 2002). Dengan menggunakan SMOTE, sampel baru disintesis dari kelas minoritas sehingga dataset menjadi lebih seimbang, dengan tujuan meningkatkan kinerja metode klasifikasi.

### 2.1.6 Ekstraksi Fitur

Ekstraksi fitur merupakan proses mengubah data teks tidak terstruktur menjadi data numerik yang terstruktur. Ekstraksi fitur penting dilakukan karena dalam algoritma *machine learning* data teks tidak dapat diproses (Budiman dkk., 2020). Salah satu metode ekstraksi fitur yang umum digunakan adalah *Term Frequency and Inverse Document frequency* (TF-IDF) yang bertujuan menilai tingkat kepentingan suatu kata dalam suatu dokumen (Riefky & Anandyani, 2019). TF mengukur frekuensi kemunculan sebuah kata dalam dokumen tertentu. Semakin sering kata tersebut muncul, maka nilai TF-nya akan semakin tinggi. Secara matematis, TF dapat dirumuskan pada persamaan (2.3) berikut:

$$TF = \frac{f(t, D)}{\text{jumlah kata dalam dokumen } (D)} \quad (2.2)$$

dimana  $f(t, D)$  adalah frekuensi kata  $t$  dalam dokumen  $D$ . IDF berfungsi mengukur tingkat unik atau pentingnya sebuah kata di dalam kumpulan dokumen, sementara DF menghitung jumlah dokumen yang memuat kata tersebut (Gifari dkk., 2022). Kata-kata yang jarang muncul dalam dokumen akan memiliki nilai

IDF yang tinggi, sementara kata-kata yang sering muncul akan memiliki nilai IDF yang lebih rendah, IDF dapat dirumuskan pada persamaan (2.4) berikut:

$$IDF = \log\left(\frac{N}{df(t)}\right) \quad (2.3)$$

dengan  $N$  adalah jumlah total dokumen, dan  $df(t)$  adalah jumlah dokumen yang memuat kata  $t$ . Secara matematis, TF-IDF merupakan hasil perkalian dari TF dan IDF, yang dirumuskan pada persamaan (2.5) berikut:

$$V = tf(t) \times idf(t) \quad (2.4)$$

di mana  $tf$  mewakili frekuensi kemunculan kata dalam dokumen, dan  $idf$  mewakili *inverse* frekuensi dalam dokumen. Kemudian hasil dari  $V$  akan dinormalisasikan dengan *Euclidian norm*. Persamaan ini berfungsi sebagai normalisasi nilai bobot yang diperoleh, supaya nilai bobot berada pada *range* yang sama, yang dirumuskan pada persamaan (2.6) berikut:

$$V_{norm} = \frac{V_n}{\sqrt{V_1^2 + V_2^2 + \dots + V_n^2}} \quad (2.5)$$

Berikut adalah contoh perhitungan TF-IDF untuk setiap kata dalam dokumen berikut:

**Tabel 2.1** Contoh Kalimat dalam Dokumen

Dokumen	Kalimat
1	Kebijakan kenaikan UKT memicu amarah publik
2	Banyak mahasiswa protes kenaikan UKT

Dokumen pada Tabel 2.1, akan dihitung frekuensi kemunculan setiap term menggunakan persamaan (2.2) Hasil perhitungan disajikan pada Tabel 2.2. Selanjutnya, nilai IDF untuk setiap term dapat dihitung menggunakan persamaan (2.3) dengan hasil disajikan pada Tabel 2.3. Nilai TF-IDF setiap kata dalam setiap dokumen diperoleh dengan mengalikan nilai TF dan IDF. Hasil perhitungan menggunakan persamaan (2.4) disajikan pada Tabel 2.4. Nilai TF-IDF yang dihasilkan merepresentasikan bobot setiap kata dalam masing-masing dokumen, di mana kata dengan bobot lebih tinggi dianggap lebih penting dalam konteks dokumen tersebut.

**Tabel 2.2** Contoh Perhitungan *Term Frequency* (TF)

Dok	Kebijakan	Banyak	Mahasiswa	Kenaikan	UKT	Memicu	Protes	Amarah	Publik
1	1	0	0	1	1	1	0	1	1
2	0	1	1	1	1	0	1	0	0

**Tabel 2.3** Contoh Perhitungan *Inverse Document Frequency* (IDF)

Kebijakan	Banyak	Mahasiswa	Kenaikan	UKT	Memicu	Protes	Amarah	Publik
0,3	0,3	0,3	0	0	0,3	0,3	0,3	0,3

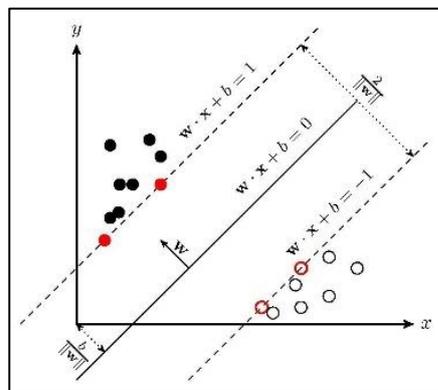
**Tabel 2.4** Contoh Perhitungan TF-IDF

Dok	Kebijakan	Banyak	Mahasiswa	Kenaikan	UKT	Memicu	Protes	Amarah	Publik
1	0,3	0	0	0	0	0,3	0	0,3	0,3
2	0	0,3	0,3	0	0	0	0,3	0	0

### 2.1.7 *Support Vector Machine* (SVM)

SVM adalah metode dalam *supervised machine learning* yang diperkenalkan oleh Vapnik tahun 1995 kemudian dikembangkan lebih lanjut oleh

Joachims pada tahun 1998. Berdasarkan Gambar 2.2, SVM digunakan untuk klasifikasi dua kelas dengan cara memisahkan data ke dalam kelas positif dan negatif menggunakan *hyperplane*, yaitu garis atau bidang yang memisahkan kedua kelas tersebut di ruang data. Ketika dataset memiliki tiga dimensi, *hyperplane* yang terbentuk adalah sebuah bidang. Namun, jika dataset memiliki lebih tinggi, SVM mencari *hyperplane* terbaik yang memaksimalkan margin dengan titik data terdekat (*Support Vector*).



**Gambar 2.2** Ilustrasi SVM

SVM digunakan untuk menyelesaikan permasalahan klasifikasi, baik yang bersifat *linear* maupun *non-linear*, sehingga menjadi populer dalam klasifikasi data (Wahyudi dkk., 2021). Formulasi optimasi SVM untuk masalah klasifikasi dibedakan menjadi dua kelas yaitu SVM *linear* dan SVM *non-linear* sebagai berikut:

### 1. SVM *Linear*

Dataset dinotasikan sebagai  $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$ , di mana setiap  $x_i \in R^n$  merepresentasikan sebuah vektor fitur berdimensi  $n$ . Indeks  $i =$

$1, 2, \dots, l$  menunjukkan urutan data, dengan  $l$  merupakan jumlah data. Setiap data  $\mathbf{x}_i$  memiliki label kelas yang dinotasikan sebagai  $y_i \in \{1, -1\}$ , dengan nilai 1 untuk kelas positif dan  $-1$  untuk kelas negatif. Persamaan *hyperplane* secara umum dapat dirumuskan pada persamaan (2.6). Sedangkan pertidaksamaan *hyperplane* kelas negatif dan kelas positif dirumuskan pada persamaan (2.7) dan persamaan (2.8), yaitu (Vapnik dkk., 1995):

$$\mathbf{w}\mathbf{x}_i + b = 0 \quad (2.6)$$

$$\mathbf{w}\mathbf{x}_i + b \geq 1, y_i = 1 \quad (2.7)$$

$$\mathbf{w}\mathbf{x}_i + b \leq -1, y_i = -1 \quad (2.8)$$

dengan  $\mathbf{w}$  adalah vektor pembobot dan  $b$  adalah bias. Jarak margin antara *hyperplane* dengan titik data terdekat dapat dihitung sebagai:

$$\text{Jarak margin} = \frac{2}{\|\mathbf{w}\|} \quad (2.9)$$

SVM mencari *hyperplane* terbaik dengan cara memaksimalkan jarak antara *hyperplane* dengan setiap kelas. Margin terbesar dapat ditemukan dengan memaksimalkan persamaan (2.9), yaitu meminimalkan nilai panjang vektor  $\mathbf{w}$ . Kemudian diformulasikan pada *Quadratic Programming* dengan meminimalkan fungsi sebagai berikut:

$$f(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.10)$$

dan memenuhi syarat persamaan (2.11) berikut:

$$y_i(\mathbf{w}\mathbf{x}_i + b) - 1 \geq 0, \forall i \quad (2.11)$$

*Quadratic Programming* memiliki beberapa metode, salah satunya ialah *Lagrange Multiplier* dengan perhitungan menggunakan persamaan berikut:

$$\begin{aligned} L &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1] \\ &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - \sum_{i=1}^l \alpha_i \end{aligned}$$

di mana  $\alpha_i$  sebagai *Lagrange Multiplier*, bernilai nol atau positif. Selanjutnya  $L$  diminimalkan terhadap  $\mathbf{w}$ ,  $b$  dan dimaksimalkan terhadap  $\alpha_i > 0$ . Dengan memperhatikan sifat bahwa pada titik optimal gradien  $L = 0$ , sebagai berikut:

$$\begin{aligned} \frac{\partial L}{\partial b} &= 0 \\ \sum_{i=1}^l \alpha_i y_i &= 0 \\ \frac{\partial L}{\partial \mathbf{w}} &= 0 \\ \mathbf{w} - \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i &= 0 \\ \mathbf{w} &= \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i \end{aligned}$$

Proses optimasi dapat dilakukan dengan cara memaksimalkan  $L$  terhadap  $\alpha_i$  melalui substitusi berikut, yaitu:

$$L = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - \sum_{i=1}^l \alpha_i$$

$$\begin{aligned}
&= \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) - \left( \sum_{i=1}^l \alpha_i y_i \mathbf{w} \mathbf{x}_i + \sum_{i=1}^l \alpha_i y_i b - \sum_{i=1}^l \alpha_i \right) \\
&= \frac{1}{2} \left( \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i + \sum_{i=1}^l \alpha_j y_j \mathbf{x}_j \right) \\
&\quad - \left( \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i + \sum_{i=1}^l \alpha_j y_j \mathbf{x}_j + 0 - \sum_{i=1}^l \alpha_i \right) \\
&= \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \mathbf{x}_i \mathbf{x}_j - \left( \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \mathbf{x}_i \mathbf{x}_j - \sum_{i=1}^l \alpha_i \right) \\
&= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i K(\mathbf{x}_i, \mathbf{x}_j) \tag{2.12}
\end{aligned}$$

di mana  $\alpha_i \geq 0, i = 1, 2, \dots, l, \sum_{i=1}^l \alpha_i y_i = 0$

Secara umum dua kelas tidak dapat terpisah sempurna dengan *hyperplane*. Oleh karena itu, *constraint* pada persamaan (2.11) tidak dapat dipenuhi, sehingga proses optimasi tidak dapat dilakukan. Untuk mengatasi masalah tersebut dapat dirumuskan ulang dengan memperkenalkan teknik *soft margin* yaitu penambahan *slack variables*  $\xi_i > 0$  pada persamaan (2.11) sebagai berikut:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \forall i \tag{2.13}$$

dengan demikian persamaan (2.10) diubah menjadi,

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i$$

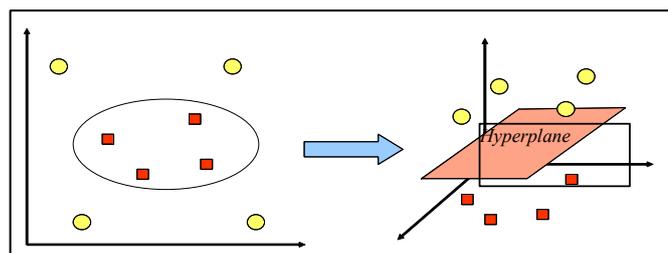
Ketika persamaan (2.13) diminimalkan, jarak margin antar kelas-kelas akan dimaksimalkan. Variabel *Slack*  $\xi_i$  berfungsi untuk mengurangi jumlah kesalahan klasifikasi (*misclassification error*). Parameter C berperan dalam

mengendalikan tingkat toleransi terhadap kesalahan klasifikasi, sehingga menjaga keseimbangan antara optimasi margin maksimum dan pengurangan kesalahan klasifikasi  $\xi$ . Nilai  $C$  yang tinggi menghasilkan model yang kompleks dengan toleransi kesalahan yang rendah, sedangkan nilai  $C$  yang rendah dapat menghasilkan model yang lebih umum dengan toleransi kesalahan yang lebih tinggi.

## 2. SVM *Non-linear*

Konsep dasar SVM melibatkan pembentukan *hyperplane* dalam ruang dimensi  $n - 1$  untuk memisahkan dua kelas pada ruang berdimensi  $n$  (Handkk., 2022). Masalah non-linier tidak memerlukan konstruksi secara eksplisit, tetapi dapat diatasi menggunakan pendekatan fungsi *kernel* (Dikkers, 2005). Fungsi *kernel* secara umum dirumuskan sebagai berikut:

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$$



**Gambar 2.3** Proses Transformasi Data dalam Ruang Dimensi Tinggi

Berdasarkan Gambar 2.3, data dari kelas positif dan kelas negatif yang berada dalam ruang dua dimensi tidak dapat dipisahkan secara *linear* menggunakan *hyperplane* (Nugroho, 2008). Proses transformasi menyelesaikan permasalahan ini dengan memetakan data dari berdimensi dua ke ruang

berdimensi lebih tinggi, misalnya dimensi tiga, sehingga kelas positif dan kelas negatif dapat dipisahkan secara *linear* oleh *hyperplane*. Proses transformasi dinotasikan sebagai berikut:

$$\Phi = R^p \rightarrow R^q, \text{ dengan } p < q$$

di mana  $p$  dan  $q$  adalah dimensi. Transformasi  $\Phi$  didefinisikan secara implisit oleh fungsi *kernel* karena transformasi ini biasanya tidak diketahui secara eksplisit. Proses menentukan klasifikasi data ditentukan melalui persamaan berikut (Vapnik dkk., 1995):

$$\begin{aligned} f(\Phi(x)) &= \text{sign}(\mathbf{w} \Phi(x) + b) \\ &= \text{sign}\left(\sum_{i=1}^n \alpha_i y_i \Phi(x_i) \Phi(x_j) + b\right) \\ &= \text{sign}\left(\sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b\right) \end{aligned} \quad (2.14)$$

di mana  $\alpha_i > 0$ . Nilai  $w$  dan  $b$  dapat dihitung dengan persamaan berikut:

$$\mathbf{w} = \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i \quad (2.15)$$

$$b = \frac{1}{NSV} \sum_{x_j \in SV} \left( \frac{1}{y_i} - \sum_{x_j \in SV} (\alpha_i y_j K(\mathbf{x}_i, \mathbf{x}_j)) \right) \quad (2.16)$$

Keterangan :

$\alpha$  : Nilai *Lagrange Multiplier*

$NSV$  : Jumlah *support vector*

$b$  : Nilai bias

$K(x_j, x_i)$ : Fungsi *Kernel*

$\text{sign}$  : Fungsi *Signum* atau fungsi tanda

Metode SVM memanfaatkan berbagai fungsi *kernel* yaitu (Yu dkk., 2010):

a. *Kernel Linear*

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \cdot \mathbf{x}_j \quad (2.17)$$

Pada *kernel linear*, parameter yang digunakan adalah  $C$ , yaitu parameter regularisasi yang mengontrol kesalahan klasifikasi pada data *training*.

b. *Kernel Polinomial*

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i \cdot \mathbf{x}_j + 1)^d$$

*Kernel* polinomial memiliki beberapa parameter penting, yaitu  $C$ , *degree* ( $d$ ), dan *gamma* ( $\gamma$ ).

c. *Kernel Sigmoid*

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i \cdot \mathbf{x}_j + C)$$

Sedangkan pada *kernel sigmoid*, parameter yang digunakan adalah  $C$  dan *gamma* ( $\gamma$ ).

d. *Kernel Radial Basis Function*

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \quad (2.18)$$

Pada *kernel* RBF, parameter utama yang digunakan adalah  $C$  dan *gamma* ( $\gamma$ ). Pemilihan kombinasi nilai  $C$  dan *gamma* sangat penting karena keduanya mempengaruhi performa klasifikasi secara signifikan.

### 2.1.8 *Grid Search dan Cross Validation*

*Grid Search* merupakan salah satu metode optimasi *hyperparameter* yang digunakan untuk mencari kombinasi parameter terbaik dari suatu model

pembelajaran mesin (Khalid & Javaid, 2020). Dalam konteks algoritma SVM, parameter seperti  $C$ ,  $\gamma$ , dan  $\text{degree}$  dapat memengaruhi performa model secara signifikan. *Grid Search* bekerja dengan mengevaluasi setiap kombinasi dari parameter yang ditentukan dalam suatu *grid* (ruang parameter). Seluruh kombinasi parameter yang mungkin kemudian diuji satu per satu, dan dipilih kombinasi dengan performa terbaik berdasarkan metrik evaluasi. Setiap kombinasi parameter tersebut kemudian dievaluasi menggunakan teknik validasi silang atau *cross-validation* (CV).

*Cross Validation* (CV) adalah teknik yang digunakan untuk mengevaluasi performa model pembelajaran mesin dengan membagi data *training* ke dalam beberapa bagian (*fold*). Salah satu metode *cross validation* yang paling umum adalah *k-fold Cross Validation*, di mana data dibagi menjadi  $k$  bagian yang sama besar. Pada setiap iterasi, satu bagian digunakan sebagai data validasi, sedangkan sisanya digunakan sebagai data *training*. Proses ini diulang sebanyak  $k$  kali, dan hasil evaluasi dirata-rata untuk mendapatkan estimasi performa model yang lebih stabil (Iriananda dkk., 2024).

Proses ini diulang untuk seluruh kombinasi parameter yang tersedia dalam *grid*. Kombinasi parameter yang menghasilkan nilai rata-rata metrik evaluasi terbaik pada proses *cross-validation* akan dipilih sebagai parameter optimal. Selanjutnya, model akhir akan dilatih ulang menggunakan seluruh data *training* dengan konfigurasi parameter terbaik tersebut, sebelum digunakan pada data *testing* atau data aktual.

### 2.1.9 Evaluasi Model

Evaluasi model klasifikasi sentimen dilakukan untuk mengukur seberapa baik model mampu mengklasifikasikan data ke dalam kategori yang tepat. *Confusion Matrix* adalah *matrix* yang digunakan untuk mengevaluasi model (Ariyanto & Chamidah, 2021). Hasil dari model di klasifikasikan berdasarkan jumlah data *testing* yang benar dan salah. Tabel klasifikasi yang digunakan adalah sebagai berikut:

**Tabel 2.5** *Confusion Matrix*

Kelas		Kelas Aktual	
		Positif	Negatif
Kelas Prediksi	Positif	TP	FP
	Negatif	FN	TN

Keterangan:

TP : Jumlah data yang benar diklasifikasikan sebagai positif

FP : Jumlah data yang salah diklasifikasikan sebagai positif

FN : Jumlah data yang salah diklasifikasikan sebagai negatif

TN : Jumlah data yang benar diklasifikasikan sebagai negatif

Berdasarkan Tabel 2.5, evaluasi model dilakukan untuk menilai efektivitas model dalam melakukan klasifikasi. Berikut adalah beberapa metrik evaluasi utama yang digunakan (Astuti, 2020):

#### 1. Akurasi

Akurasi adalah ukuran yang menunjukkan seberapa baik model dalam memprediksi secara keseluruhan. Rumus untuk menghitung Akurasi adalah:

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (2.19)$$

## 2. *Presisi*

Akurasi hanya memberikan gambaran umum mengenai performa model, namun tidak selalu mencerminkan performa yang memadai terutama dalam kasus ketidakseimbangan kelas pada data. Oleh karena itu, presisi digunakan untuk mengukur seberapa banyak prediksi positif yang benar-benar tepat. Presisi memberikan informasi tentang seberapa banyak prediksi positif yang benar dari semua prediksi positif yang dibuat oleh model. Rumus untuk menghitung nilai presisi adalah:

$$Presisi = \frac{TP}{TP + FP} \quad (2.20)$$

## 3. *Recall* atau sensitivitas

*Recall* mengukur seberapa banyak dari nilai aktual yang diprediksi dengan benar oleh model. Semakin tinggi nilai *recall*, semakin baik model dalam mendeteksi seluruh data yang sebenarnya positif. Rumus menghitung nilai *recall*:

$$Recall = \frac{TP}{TP + FN} \quad (2.21)$$

## 4. *F1-Score*

*F1-Score* merupakan nilai *mean* dari presisi dan *recall*. Rumus menghitung nilai *F1-Score*:

$$F1 - score = \frac{2 \cdot Presisi \cdot Recall}{Presisi + Recall} \quad (2.22)$$

5. ROC (*Receiver Operating Characteristic*) dan AUC (*Area Under the Curve*).

ROC merupakan grafik yang digunakan untuk mengevaluasi kinerja model klasifikasi, terutama pada permasalahan klasifikasi biner (Roifa, 2018). Grafik ini menggambarkan hubungan antara *precision* dan *recall*, dengan memplotkan *True Positive Rate* (TPR) pada sumbu y dan *False Positive Rate* (FPR) pada sumbu x. Grafik ROC memiliki keunggulan dalam memvisualisasikan serta mengatur kinerja *classifier* tanpa bergantung pada distribusi kelas atau biaya kesalahan (*error cost*). TPR menggambarkan proporsi prediksi kelas positif yang benar dibandingkan dengan kondisi aktual. Sementara itu, FPR menunjukkan jumlah kesalahan prediksi kelas positif ketika sebenarnya kelas tersebut negatif. Model dikatakan baik apabila memiliki nilai TPR yang tinggi dan FPR yang rendah. Nilai TPR dan FPR dihitung berdasarkan *confusion matrix*.

Model klasifikasi memprediksi data berdasarkan peluang tiap kelas dengan menghasilkan kurva ROC dari pasangan koordinat FPR dan TPR, yang dihitung berdasarkan *threshold* tertentu. FP dan TP diperoleh dari peluang kesalahan dan peluang kebenaran yang melebihi ambang batas, lalu dinormalisasi terhadap total data. Koordinat (FPR,TPR) dihubungkan membentuk kurva ROC dengan rentang nilai antara 0 dan 1. Setiap perubahan *threshold* menghasilkan titik baru pada ruang ROC. Kinerja model diukur melalui AUC, yakni luas area di bawah kurva ROC.

Perhitungan nilai TPR menggunakan persamaan (2.23) dan FPR menggunakan persamaan (2.24) berikut:

$$TPR = \frac{TP}{TP + FN} \quad (2.23)$$

$$\begin{aligned} FPR &= 1 - \text{specificity} \\ &= 1 - \frac{TN}{FP + TN} \end{aligned} \quad (2.24)$$

Nilai AUC dapat dirumuskan pada persamaan (2.25) berikut:

$$AUC = \frac{1}{2} \sum_{i=1}^l (TPR_{i+1} + TPR_i)(FPR_{i+1} - FPR_i) \quad (2.25)$$

Interpretasi nilai AUC disajikan dalam Tabel 2.6.

**Tabel 2.6** Nilai AUC dan Interpretasi

Nilai AUC	Interpretasi
0,9 – 1	Klasifikasi sangat baik
0,8 - 0,9	Klasifikasi baik
0,7 – 0,8	Klasifikasi cukup
0,6 - 0,7	Klasifikasi lemah
0,5 – 0,6	Klasifikasi gagal

#### 2.1.10 Uang Kuliah Tunggal (UKT)

Uang Kuliah Tunggal (UKT) adalah sistem penetapan biaya kuliah yang diperkenalkan oleh pemerintah Indonesia dengan tujuan untuk menciptakan keadilan dalam pembebanan biaya pendidikan tinggi kepada mahasiswa. Sistem ini menetapkan bahwa setiap mahasiswa di Perguruan Tinggi Negeri (PTN) membayar sejumlah biaya yang ditentukan berdasarkan kemampuan ekonomi keluarganya. Dengan demikian, mahasiswa dari latar belakang ekonomi yang

kurang mampu dapat membayar biaya kuliah yang lebih rendah dibandingkan dengan mahasiswa yang berasal dari keluarga dengan kemampuan ekonomi lebih tinggi, sehingga pendidikan tinggi menjadi lebih terjangkau dan merata (Suyoga dkk., 2017).

Pemerintah Indonesia melalui Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia No. 39 Tahun 2017 tentang Biaya Kuliah Tunggal dan Uang Kuliah Tunggal pada Perguruan Tinggi Negeri di Lingkungan Kementerian Riset, Teknologi, dan Pendidikan Tinggi, mengatur tentang mekanisme penetapan UKT. Peraturan ini menetapkan bahwa UKT adalah sebagian dari BKT yang harus ditanggung oleh mahasiswa setiap semester.

#### **2.1.11 Platform X**

Platform X merupakan salah satu media sosial yang berfungsi sebagai wadah untuk pertukaran informasi dan interaksi antar pengguna (Juwiantho dkk., 2020). Di Platform X, pengguna dapat berbagi pesan yang dikenal sebagai *tweet*, yang dapat berupa teks, foto, video, atau tautan. *Tweet* yang diunggah dapat dilihat di profil pengguna, dikirim kepada pengikutnya, dan dicari melalui fitur pencarian Platform X. Setiap *tweet* dilengkapi dengan fitur-fitur seperti *Retweet*, yang memungkinkan pengguna untuk membagikan kembali *tweet* kepada pengikut mereka, serta fitur balasan untuk memberikan tanggapan publik terhadap *tweet* orang lain. Fitur balasan ini memungkinkan semua pengguna melihat tanggapan tersebut, berbeda dengan *Direct Message* yang bersifat pribadi dan hanya terlihat oleh pengirim dan penerima pesan. Selain itu, terdapat fitur *Like* yang memungkinkan pengguna untuk menunjukkan bahwa mereka menyukai sebuah

*tweet*, serta fitur *Share* yang memungkinkan pengguna untuk membagikan *tweet* melalui *Direct Message*, menyalin tautan *tweet*, atau menyimpannya di daftar *Bookmark*. Platform X juga dikenal karena sering memuat topik-topik terkini dan peristiwa-peristiwa yang sedang hangat dibahas oleh pengguna.

Platform X merupakan media sosial yang sebelumnya dikenal dengan nama Twitter. Pada tahun 2023, Twitter mengalami perubahan nama dan identitas visual menjadi X, sebagai bagian dari *rebranding* besar yang dilakukan oleh pemiliknya, Elon Musk. Meskipun terjadi perubahan nama, fungsi dan peran platform ini sebagai media sosial yang digunakan secara luas oleh masyarakat tetap sama.

## 2.2 Kajian Integrasi Topik dengan Al-Quran/Hadits

Al-Quran merupakan sumber utama ajaran Islam yang memberikan pedoman hidup bagi umat manusia dalam berbagai aspek kehidupan, termasuk dalam pengambilan keputusan kebijakan publik dan penyampaian aspirasi masyarakat. Dalam hal kebijakan kenaikan UKT, terdapat beberapa nilai Islam yang dapat dijadikan landasan evaluatif, yakni pentingnya pertimbangan objektif, musyawarah, dan menjaga etika komunikasi. Islam menekankan pentingnya musyawarah dalam mengambil keputusan yang menyangkut kepentingan publik. Hal ini ditegaskan dalam QS. Asy-Syura ayat 38:

وَالَّذِينَ اسْتَجَابُوا لِرَبِّهِمْ وَأَقَامُوا الصَّلَاةَ وَأَمْرُهُمْ شُورَىٰ بَيْنَهُمْ وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ ﴿٣٨﴾

(Kementerian Agama RI, 2022)

Artinya : “Dan (bagi) orang-orang yang menerima (mematuhi) seruan Tuhannya dan mendirikan shalat, sedang urusan mereka (diputuskan) dengan musyawarah

*antara mereka; dan mereka menafkahkan sebagian dari rezeki yang Kami berikan kepada mereka.”*

Tafsir Ibnu Katsir menjelaskan bahwa ayat ini menunjukkan keutamaan bersikap syura (musyawarah) dalam memutuskan segala urusan bersama. Dalam konteks kebijakan kenaikan UKT, ayat ini dapat diinterpretasikan sebagai landasan bahwa proses pengambilan kebijakan seharusnya tidak dilakukan secara sepihak oleh otoritas kampus atau pemerintah, melainkan melalui proses dialog yang melibatkan seluruh pemangku kepentingan, seperti mahasiswa, orang tua, akademisi, dan masyarakat luas (Ghoffar dkk., 2004).

Selain musyawarah, Islam juga mengajarkan etika dalam menyampaikan aspirasi dan kritik terhadap kebijakan publik. Etika ini tercermin dalam Al-Qur'an Surah Al-Hujurat ayat 11, di mana Allah SWT berfirman:

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا يَسْخَرْ قَوْمٌ مِّن قَوْمٍ عَسَىٰ أَن يَكُونُوا خَيْرًا مِّنْهُمْ وَلَا نِسَاءٌ مِّن نِّسَاءٍ عَسَىٰ  
 أَن يَكُنَّ خَيْرًا مِّنْهُنَّ وَلَا تَلْمِزُوا أَنفُسَكُمْ وَلَا تَنَابَزُوا بِالْأَلْقَابِ بِئْسَ الْأِسْمُ الْفُسُوقُ بَعْدَ الْإِيمَانِ  
 وَمَنْ لَّمْ يَتُبْ فَأُولَٰئِكَ هُمُ الظَّالِمُونَ ﴿١١﴾

(Kementerian Agama RI, 2022)

Artinya : *“Wahai orang-orang yang beriman, janganlah suatu kaum mengolok-olok kaum yang lain (karena) boleh jadi mereka (yang diolok-olokkan itu) lebih baik daripada mereka (yang mengolok-olok) dan jangan pula perempuan-perempuan (mengolok-olok) perempuan lain (karena) boleh jadi perempuan (yang diolok-olok itu) lebih baik daripada perempuan (yang mengolok-olok). Janganlah kamu saling mencela dan saling memanggil dengan julukan yang buruk. Seburuk-buruk panggilan adalah (panggilan) fasik setelah beriman. Siapa yang tidak bertobat, mereka itulah orang-orang zalim.”*

Menurut Ibnu Katsir, ayat ini dimaknai sebagai larangan tegas terhadap segala bentuk penghinaan, cercaan, serta tindakan merendahkan martabat orang lain

(Ghoffar dkk., 2004). Hal ini tidak hanya berlaku dalam komunikasi langsung, tetapi juga dalam bentuk komunikasi tidak langsung seperti tulisan, opini publik, serta komentar di media sosial. Penekanan pada pentingnya menjaga kehormatan sesama merupakan prinsip utama dalam menjaga kesatuan dan harmoni sosial.

Dalam konteks era digital saat ini, penyampaian kritik terhadap kebijakan seperti kenaikan UKT sering disuarakan melalui platform media sosial. Oleh karena itu, ayat ini relevan sebagai pedoman moral bahwa penyampaian kritik hendaknya dilakukan secara santun, objektif, dan tidak mengandung unsur kebencian, fitnah, ataupun provokasi. Kebebasan berpendapat dalam Islam tidak berarti kebebasan tanpa batas, melainkan kebebasan yang diimbangi dengan tanggung jawab dan etika.

### **2.3 Kajian Topik dengan Teori Pendukung**

Penelitian analisis sentimen terhadap kebijakan kenaikan UKT di platform X menggunakan metode SVM melibatkan serangkaian tahapan yang didukung oleh teori-teori yang relevan. Tahapan penelitian ini dimulai dengan pengumpulan data *tweet* yang berkaitan dengan kebijakan kenaikan UKT. Pengambilan data dilakukan dengan proses web *scraping* dengan memasukkan *query* “UKT” pada fitur pencarian di halaman platform X. Selanjutnya, data tersebut melalui tahap *preprocessing* yang melibatkan langkah-langkah seperti *cleaning*, *tokenizing*, *case folding*, *stopword removal*, dan *stemming*. Langkah-langkah *preprocessing* ini mempersiapkan data teks sebagai *input* untuk model SVM.

Setelah *preprocessing*, dilakukan visualisasi data menggunakan *word cloud*. Tahap berikutnya adalah pelabelan data, yang dilakukan secara manual oleh peneliti

dengan mengkategorikan *tweet* ke dalam sentimen positif dan negatif. Pelabelan dilakukan berdasarkan kriteria yang telah ditetapkan dalam penelitian sebelumnya (Bahri dkk., 2018).

Tahap berikutnya adalah pembagian data, di mana dataset dibagi menjadi dua bagian, yaitu data *training* dan data *testing*. Pembagian data dilakukan secara acak untuk memastikan keseimbangan dalam analisis model. Setelah data dibagi, dilakukan klasifikasi menggunakan algoritma SVM. Pada tahap ini, data *training* digunakan untuk membangun model klasifikasi sentimen. Sebelum proses pelatihan, dilakukan perhitungan bobot kata menggunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik. Selain itu, dilakukan teknik SMOTE untuk menangani ketidakseimbangan jumlah data. Setelah proses tersebut, model dilatih menggunakan algoritma SVM untuk mengenali pola sentimen dalam data *training*. Pelatihan dilakukan menggunakan SVM dengan berbagai jenis *kernel*, yaitu *linear*, *polynomial*, *radial basis function (RBF)*, dan *sigmoid*. Masing-masing kernel diuji dengan kombinasi parameter yang berbeda untuk memperoleh konfigurasi terbaik. Proses pencarian parameter terbaik dilakukan melalui pendekatan *trial and error* dan dioptimalkan menggunakan metode *Grid Search* yang dilengkapi dengan *Cross-Validation*.

Selanjutnya, model yang telah dilatih dievaluasi menggunakan data *testing*. Sebelum proses klasifikasi, data *testing* juga melalui transformasi TF-IDF agar memiliki bentuk representasi yang sama dengan data pelatihan. Namun, teknik SMOTE tidak diterapkan pada data *testing*, karena metode tersebut hanya digunakan pada data pelatihan untuk membentuk distribusi yang seimbang. Data *testing* kemudian dimasukkan ke dalam model SVM untuk diprediksi label

sentimennya berdasarkan pola yang telah dipelajari saat pelatihan. Beberapa metrik evaluasi digunakan untuk mengukur kinerja model, seperti *confusion matrix* yang menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas, serta akurasi, presisi, *recall*, *F1-Score* dan ROC AUC. untuk memberikan gambaran yang lebih komprehensif mengenai efektivitas model dalam mengklasifikasikan sentimen *tweet*.

Analisis hasil dari prediksi model membantu dalam memahami interpretasi dari sentimen yang diberikan terhadap kebijakan kenaikan UKT. Dengan memadukan teori-teori ini, penelitian dapat menambah wawasan yang lebih mendalam tentang evaluasi sentimen publik terhadap kebijakan kenaikan UKT menggunakan pendekatan SVM di platform X.

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Jenis Penelitian**

Jenis Penelitian ini merupakan penelitian kuantitatif. Penelitian kuantitatif melibatkan pengumpulan dan analisis data dalam bentuk numerik untuk mengidentifikasi pola dan tren dalam sentimen *tweet*. Data yang diperoleh akan dianalisis secara statistik untuk menghasilkan kesimpulan yang terukur.

#### **3.2 Data dan Sumber Data**

Data yang digunakan dalam penelitian ini adalah kumpulan *tweet* yang berkaitan dengan kebijakan kenaikan UKT. Sumber data yang digunakan bersumber dari Platform X menggunakan kata kunci “UKT” pada *tweet* berbahasa Indonesia dengan rentang waktu dari 1 Januari 2024 hingga 25 Mei 2024.

#### **3.3 Teknik Pengumpulan Data**

Data dikumpulkan dengan menggunakan teknik web *scraping* melalui API (*Application Programming Interface*) yang didapatkan dari Platform X. Web *scraping* adalah proses otomatisasi untuk mengekstrak data dari situs web yang dilakukan dengan kode *python* di Google Colab. Proses ini melibatkan pencarian dan pengumpulan *tweet* yang berkaitan dengan kebijakan kenaikan UKT dengan menggunakan kata kunci “UKT”.

### 3.4 Teknik Analisis Data

Adapun langkah-langkah yang dilakukan dalam penelitian adalah sebagai berikut:

#### 1. Pengumpulan data

Pada tahap ini, pengumpulan data dilakukan dengan mengambil 1000 *tweet* yang berkaitan dengan kenaikan Uang Kuliah Tunggal (UKT). Pengumpulan data dilakukan dengan teknik *web scraping* melalui API Platform X. *Web scraping* dilakukan dengan menggunakan kata kunci "UKT" pada *tweet* berbahasa Indonesia dengan rentang waktu dari 1 Januari 2024 hingga 25 Mei 2024.

#### 2. Pelabelan data

Proses pelabelan data dilakukan dengan menandai data sesuai dengan kategori tertentu. Pada tahap ini, *tweet* yang diperoleh dari Platform X diklasifikasikan ke dalam dua kategori, yaitu sentimen positif dan negatif. Pelabelan dilakukan secara manual oleh peneliti. Seluruh *tweet* akan diberi label kelas secara manual dengan mengacu pada kriteria yang tercantum dalam Tabel 3.1 (Bahri dkk., 2018)

**Tabel 3.1** Kriteria Pelabelan Data

<b>Kelas Positif</b>	<b>Kelas Negatif</b>
Suatu kalimat di kategorikan sebagai sentimen positif jika memiliki kalimat yang mengandung optimisme, prasangka baik, unsur doa, motivasi, ajakan menuju hal yang baik, mengekspresikan rasa senang dan dukungan terhadap kenaikan UKT.	Suatu kalimat di kategorikan sebagai sentimen negatif jika mengandung prasangka buruk, olok-an, hasutan, pesimisme, ketidakpercayaan, penghinaan, kritik, ekspresi kekecewaan, ketidaksenangan, keluhan, dan ujaran yang merendahkan pihak tertentu akibat kenaikan UKT.

### 3. *Preprocessing data*

Pada tahap ini, dilakukan beberapa langkah untuk mempersiapkan data agar siap digunakan dalam proses pelatihan model:

- a. *Case folding*: Semua huruf dalam teks diubah menjadi huruf kecil, untuk memastikan data memiliki struktur yang sama. Hasilnya disimpan dalam kolom baru dengan nama *casefolding\_text*.
- b. *Cleaning text*: Pada tahap ini dilakukan pembersihan data, menghapus karakter yang tidak dibutuhkan seperti emoji, tautan, hashtag, tanda baca, angka. Hasilnya disimpan dalam kolom *cleaning\_text*.
- c. *Normalisasi* : Pada tahap ini dilakukan perubahan kalimat yang tidak baku atau *slangword* menjadi kalimat baku yang sesuai dengan KBBI (Kamus Besar Bahasa Indonesia).
- d. *Tokenizing*: Setiap kalimat dipisah menjadi kata per kata atau token. Hasil tokenisasi disimpan dalam kolom baru dengan nama *tokenizing\_text*.
- e. *Stopword Removal*: Kata-kata umum yang tidak memberikan kontribusi dihapus, sehingga mengurangi jumlah kata yang disimpan oleh sistem, seperti “dan”, “atau”, dan “di”. Hasilnya disimpan dalam kolom baru dengan nama *stopword\_text*.
- f. *Stemming*: Kata-kata yang telah melalui proses *stopword removal* dikembalikan ke bentuk dasarnya sesuai dengan aturan KBBI. Proses *stemming* ini akan menggunakan *library* Sastrawi dalam *Python* yang sudah di *install* di awal. Hasilnya disimpan dalam kolom *stemming\_text*.

#### 4. Visualisasi data

Data divisualisasikan menggunakan *word cloud* untuk menampilkan kata-kata yang sering muncul. Frekuensi kemunculan data ditunjukkan oleh ukuran huruf, di mana semakin besar ukuran kata menunjukkan semakin sering kata tersebut muncul dalam dokumen.

#### 5. Pembagian Data

Setelah melalui tahap *preprocessing*, data dibagi menjadi dua kelompok, yaitu data *training* dan data *testing*.

#### 6. Tahap klasifikasi menggunakan algoritma SVM

##### a. Pelatihan Model

Pelatihan model dilakukan menggunakan data *training*. Sebelum proses klasifikasi, dilakukan perhitungan bobot kata menggunakan TF-IDF untuk mengubah data teks menjadi numerik. Selanjutnya, karena distribusi data tidak seimbang, diterapkan teknik SMOTE untuk menangani ketidakseimbangan jumlah data antara kategori sentimen positif dan negatif. Proses pelatihan dilakukan dengan menerapkan SVM menggunakan berbagai jenis *kernel*, yaitu, *linear*, polinomial, RBF dan *sigmoid*. Setiap *kernel* diuji dengan berbagai kombinasi parameter untuk menentukan konfigurasi terbaik. Untuk memperoleh kombinasi parameter yang optimal, dilakukan pencarian parameter menggunakan metode *trial and error* (Al Azies dkk., 2019). Pemilihan parameter terbaik dilakukan dengan metode *Grid Search* yang dilengkapi dengan *Cross-Validation*. Hasil dari proses pelatihan model adalah model klasifikasi yang telah dilatih dan dioptimasi dengan *kernel* serta parameter terbaik.

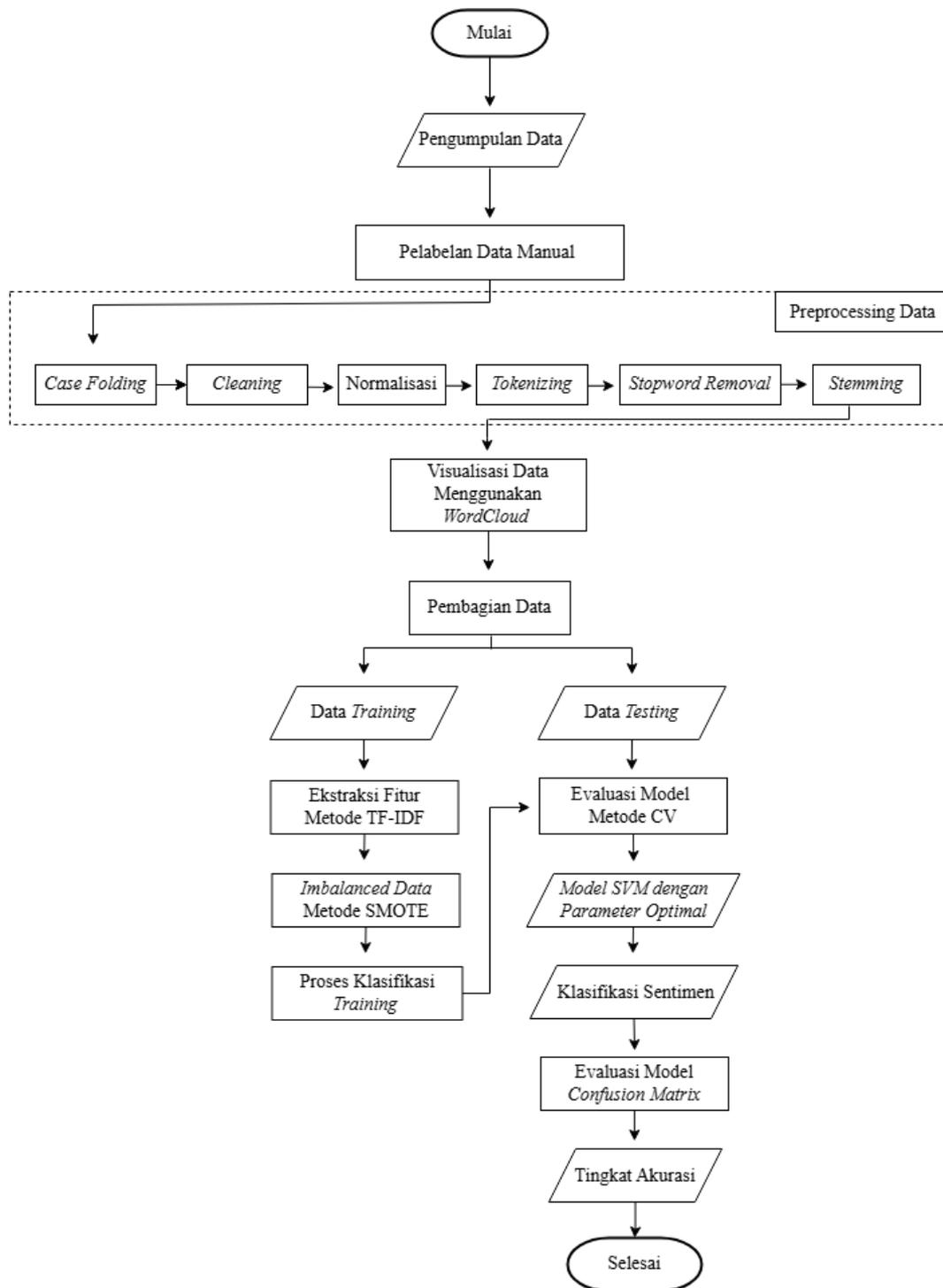
## b. Evaluasi Model

Model yang telah dilatih selanjutnya dievaluasi menggunakan data *testing*. Sebelum diklasifikasikan, data *testing* juga diolah menggunakan TF-IDF. Namun, proses SMOTE tidak diterapkan pada data *testing*, karena SMOTE hanya digunakan pada tahap pelatihan. Proses *testing* dimulai dengan memasukkan data *testing* yang telah melalui ekstraksi fitur ke dalam model SVM yang telah dilatih sebelumnya. Model kemudian memprediksi label sentimen dari data *testing* berdasarkan pola yang telah dipelajari. Hasil prediksi tersebut dievaluasi menggunakan beberapa metrik evaluasi, yaitu akurasi, presisi, *recall*, *F1-score* dan ROC AUC. Evaluasi dilakukan untuk mengetahui seberapa baik model dapat melakukan generalisasi terhadap data baru, serta untuk menilai kinerja model yang telah dilatih dalam kondisi nyata.

## 7. Interpretasi Hasil

Hasil evaluasi dianalisis untuk menilai efektivitas metode klasifikasi yang diterapkan. Interpretasi hasil mencakup penjelasan mengenai kinerja model berdasarkan metrik evaluasi yang digunakan.

### 3.2 Diagram Alir Penelitian



**Gambar 3.1** Diagram Alir Penelitian

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Pengumpulan Data

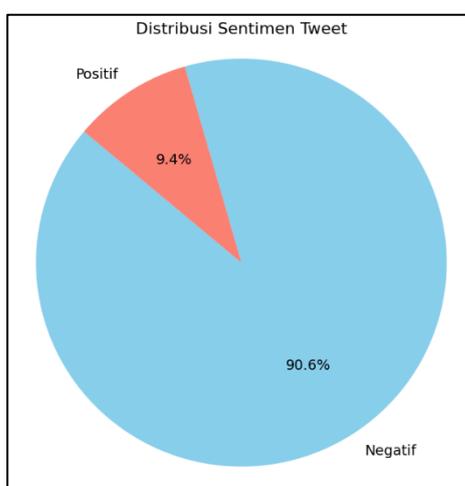
Pengumpulan data dalam penelitian ini dilakukan dengan menerapkan teknik *web scraping*, yaitu proses pengambilan data secara otomatis dari platform media sosial berdasarkan kata kunci tertentu yang telah ditentukan sebelumnya. Data yang dikumpulkan berasal dari platform X yang membahas kebijakan kenaikan UKT. Proses *web scraping* dilakukan menggunakan bahasa pemrograman *Python* melalui *Google Collaboratory*, *script data crawling* dapat dilihat pada lampiran 1.

Akses data dari platform X diperoleh melalui autentikasi API resmi yang disediakan oleh platform X. Autentikasi dilakukan dengan menggunakan *access token* yang berasal dari akun pribadi peneliti. Pengumpulan data dilakukan berdasarkan kata kunci yang relevan dan dalam rentang waktu yang telah ditetapkan.

Proses pengambilan data dilakukan secara bertahap dan berulang, menyesuaikan dengan batas maksimum permintaan (*rate limit*) data yang diberlakukan oleh platform X. Kata kunci yang digunakan akan menarik semua *tweet* yang mengandung kata tersebut, termasuk *tweet* berupa berita atau iklan yang tidak mencerminkan opini pengguna. Oleh karena itu, dilakukan proses *filtering* dengan mengecualikan *tweet* yang mengandung tautan dan balasan, agar data yang diperoleh merupakan opini murni dari pengguna. Hasil dari proses ini berupa 1.003 *tweet* yang sesuai dengan topik penelitian.

## 4.2 Pelabelan Data

Setelah data terkumpul, dilakukan proses pelabelan secara manual. Setiap *tweet* dianalisis dan dikategorikan ke dalam dua jenis sentimen, yaitu sentimen positif dan sentimen negatif. Berdasarkan hasil pelabelan, diperoleh 899 *tweet* mengandung sentimen negatif dan 103 *tweet* mengandung sentimen positif. Data yang tidak memenuhi kriteria pada Tabel 3.1 tidak dilibatkan dalam analisis lanjutan.



**Gambar 4.1** Perbandingan Jumlah *Tweet* Berdasarkan Kategori Sentimen

Gambar 4.1, menunjukkan distribusi sentimen publik terhadap kebijakan kenaikan UKT. Diagram tersebut menggambarkan perbandingan jumlah *tweet* dengan sentimen negatif dan positif, di mana jumlah *tweet* dengan sentimen negatif jauh lebih dominan dibandingkan dengan sentimen positif. Contoh data yang telah melalui proses analisis dan klasifikasi berdasarkan kategori sentimen disajikan pada Tabel 4.1.

**Tabel 4.1** Sampel Data *Tweet*

<i>Tweet</i>	Label
pak nadiem tolong dong jangan naik kan ukt kuliah orang tua saya ribut terus setiap pagi karna pendidikan yang semakin mahal sedangkan pendapatan yang ga seberapa ini harus menghidupi 4 manusia dirumah saya jadi bimbang harus kuliah atau enggak tapi kerja harus minimal sarjana	Negatif
Kalau UKT naik untuk mahasiswa baru ya gak bakal ada yg mau kuliah lagi lah kocak si Nadiem	Negatif
Setuju ukt naik asal kesejahteraan segala pendidik dan tenaga kampus juga naik	Positif
sesedih itu aku ngeliat ukt pada naik drastis	Negatif
sedih bgt UKT naik nya gila gilaan	Negatif

### 4.3 Text Preprocessing

*Preprocessing* adalah tahap awal dalam persiapan data untuk pengolahan data klasifikasi dari sebuah data yang tidak terstruktur menjadi lebih terstruktur. Pada klasifikasi teks, tahap *preprocessing* bertujuan untuk meningkatkan akurasi klasifikasi data. *Preprocessing* dilakukan dengan menggunakan *library* sastrawi yang biasa digunakan dalam NLP dalam bahasa Indonesia, *script* data *preprocessing* terlampir pada Lampiran 2. Hasil *preprocessing* terlampir pada lampiran 8 dan hasil *preprocessing* untuk sampel data sebagai berikut:

#### 1. Case folding

Pada tahap ini, semua huruf dalam data *tweet* yang terkumpul dalam data frame “*tweet\_df*” diubah menjadi huruf kecil. *Case folding* dilakukan agar proses analisis selanjutnya maksimal. Pada Tabel 4.2 dapat dilihat hasil dari sebelum dan sesudah proses *Case Folding* pada data *tweet*.

#### 2. Cleaning Text

Setelah proses *case folding* selesai, tahap selanjutnya adalah proses *cleaning* yang bertujuan untuk menghilangkan karakter yang tidak penting dalam sebuah

kalimat. Langkah ini dilakukan agar teks menjadi lebih bersih sehingga dapat mempermudah mesin dalam melakukan proses klasifikasi. Selain itu, data duplikat juga dihapus pada tahap ini untuk menjaga kualitas dan akurasi model. Dari total 1.003 data awal yang terkumpul, terdapat 9 data yang terdeteksi sebagai duplikat. Setelah penghapusan, jumlah data yang tersedia untuk analisis menjadi 985 data. Hasil dari proses *cleaning text* dapat dilihat pada Tabel 4.3.

**Tabel 4.2** Hasil Proses *Case Folding Data Tweet*

<i>Tweet</i>	<b>lower_text</b>
mau kritik ttg naik nya ukt tpi kaya percuma ga sih wkwk toh orang orang nya pd kapitalis.... klo dibatalin ttg peraturan ukt ya berarti mereka lg pertahanin jabatan nya bukan mikirin sdm anak bangsa kedepan nya #cacatberfikir	mau kritik ttg naik nya ukt tpi kaya percuma ga sih wkwk toh orang orang nya pd kapitalis.... klo dibatalin ttg peraturan ukt ya berarti mereka lg pertahanin jabatan nya bukan mikirin sdm anak bangsa kedepan nya #cacatberfikir

**Tabel 4.3** Hasil Proses *Cleaning Data Tweet*

<b>lower_text</b>	<b>cleaning_text</b>
mau kritik ttg naik nya ukt tpi kaya percuma ga sih wkwk toh orang orang nya pd kapitalis.... klo dibatalin ttg peraturan ukt ya berarti mereka lg pertahanin jabatan nya bukan mikirin sdm anak bangsa kedepan nya #cacatberfikir	mau kritik ttg naik nya ukt tpi kaya percuma ga sih wkwk toh orang orang nya pd kapitalis klo dibatalin ttg peraturan ukt ya berarti mereka lg pertahanin jabatan nya bukan mikirin sdm anak bangsa kedepan nya cacatberfikir

### 3. Normalisasi

Pada tahap ini, dilakukan proses normalisasi yang bertujuan untuk mengubah kata slang atau tidak baku menjadi bentuk baku. Proses normalisasi dilakukan dengan menggunakan kamus kata normalisasi yang berisi pasangan kata tidak

baku dan baku yang terlampir pada Lampiran 5. Hasil dari proses normalisasi dapat dilihat pada Tabel 4.4.

**Tabel 4.4** Hasil Proses Normalisasi Data *Tweet*

<b>cleaning_text</b>	<b>slangs_text</b>
mau kritik ttg naik nya ukt tpi kaya percuma ga sih wkwk toh orang orang nya pd kapitalis klo dibatalin ttg peraturan ukt ya berarti mereka lg pertahanin jabatan nya bukan mikirin sdm anak bangsa kedepan nya cacatberfikir	mau kritik tentang naik nya ukt tapi kaya percuma tidak sih wkwk kan orang orang nya pada kapitalistik kalau dibatalkan tentang peraturan ukt ya berarti mereka lagi pertahanin jabatan nya bukan mikirin sumber daya manusia anak bangsa kedepan nya cacatberfikir

#### 4. *Tokenizing*

Pada tahap ini, dilakukan proses tokenisasi yang bertujuan untuk membagi kalimat, paragraf, atau dokumen menjadi bagian-bagian yang lebih kecil, yaitu token atau kata-kata terpisah. Hasil dari proses *tokenizing* dapat dilihat pada Tabel 4.5.

**Tabel 4.5** Hasil Proses Tokenizing Data *Tweet*

<b>slangs_text</b>	<b>token_text</b>
mau kritik tentang naik nya ukt tapi kaya percuma tidak sih wkwk kan orang orang nya pada kapitalistik kalau dibatalkan tentang peraturan ukt ya berarti mereka lagi pertahanin jabatan nya bukan mikirin sumber daya manusia anak bangsa kedepan nya cacatberfikir	['mau', 'kritik', 'tentang', 'naik', 'nya', 'ukt', 'tapi', 'kaya', 'percuma', 'tidak', 'sih', 'wkwk', 'kan', 'orang', 'orang', 'nya', 'pada', 'kapitalistik', 'kalau', 'dibatalkan', 'tentang', 'peraturan', 'ukt', 'ya', 'berarti', 'mereka', 'lagi', 'pertahanin', 'jabatan', 'nya', 'bukan', 'mikirin', 'sumber', 'daya', 'manusia', 'anak', 'bangsa', 'kedepan', 'nya', 'cacatberfikir']

### 5. Stopwords Removal

Setelah proses tokenisasi, dilakukan tahap *stopword removal* untuk menghapus kata-kata yang tidak memiliki makna signifikan dalam analisis teks. Proses ini dibantu oleh *library Sastrawi*, yang menyediakan daftar kata *stopword* dalam bahasa Indonesia. Kamus *stopword text* terlampir pada Lampiran 6. Hasil dari proses *stopwords* dapat dilihat pada Tabel 4.6.

**Tabel 4.6** Hasil Proses *Stopwords Removal* Data *Tweet*

<b>token_text</b>	<b>stopword_text</b>
['mau', 'kritik', 'tentang', 'naik', 'nya', 'ukt', 'tapi', 'kaya', 'percuma', 'tidak', 'sih', 'wkwk', 'kan', 'orang', 'orang', 'nya', 'pada', 'kapitalistik', 'kalau', 'dibatalkan', 'tentang', 'peraturan', 'ukt', 'ya', 'berarti', 'mereka', 'lagi', 'pertahananin', 'jabatan', 'nya', 'bukan', 'mikiran', 'sumber', 'daya', 'manusia', 'anak', 'bangsa', 'kedepan', 'nya', 'cacatberfikir']	['kritik', 'naik', 'ukt', 'kaya', 'tidak', 'orang', 'kapitalistik', 'dibatalkan', 'peraturan', 'ukt', 'pertahananin', 'jabatan', 'mikiran', 'sumber', 'daya', 'manusia', 'anak', 'bangsa', 'kedepan', 'cacatberfikir']

### 6. Stemming

Proses *stemming* dilakukan untuk mengubah kata berimbuhan menjadi bentuk dasarnya. Tahap ini bertujuan untuk menyederhanakan kata sehingga lebih mudah dianalisis. Hasil dari proses *Stemminng* dapat dilihat pada Tabel 4.7.

**Tabel 4.7** Hasil Proses *Stemminng* Data *Tweet*

<b>stopword_text</b>	<b>stemmed_text</b>
['kritik', 'naik', 'ukt', 'kaya', 'tidak', 'orang', 'kapitalistik', 'dibatalkan', 'peraturan', 'ukt', 'pertahananin', 'jabatan', 'mikiran', 'sumber', 'daya', 'manusia', 'anak', 'bangsa', 'kedepan', 'cacatberfikir']	kritik naik ukt kaya tidak orang kapitalistik batal atur ukt pertahananin jabat mikir sumber daya manusia anak bangsa depan cacatberfikir



#### 4.5 Pembagian Data

Pada tahap ini, data dibagi menjadi dua kelompok, yaitu data *training* dan data *testing*, dengan perbandingan 80% untuk *training* dan 20% untuk *testing*. Pembagian ini bertujuan untuk memastikan bahwa model dapat mempelajari pola sentimen publik terhadap kebijakan kenaikan UKT secara optimal, sekaligus menguji kemampuan model dalam mengklasifikasikan sentimen pada data baru.

**Tabel 4.8** Akurasi Perbandingan Data *Training* dan Data *Testing*

<b>Data Training</b>	<b>Data Testing</b>	<b>Akurasi</b>
90%	10%	0,929293
80%	20%	0,9887
70%	30%	0,918919

Pemilihan rasio 80:20 ini didasarkan pada hasil percobaan awal dengan beberapa variasi rasio pembagian, di antaranya 70:30, 80:20, dan 90:10 pada Tabel 4.8. Setiap rasio diuji dengan menerapkan tahapan klasifikasi yang sama untuk mengevaluasi performa model. Hasil evaluasi menunjukkan bahwa pembagian dengan rasio 80:20 memberikan akurasi dan kestabilan model yang paling optimal. Rasio ini memungkinkan model memperoleh cukup banyak data untuk proses pelatihan sehingga mampu mengenali pola sentimen secara efektif, sekaligus menyediakan data *testing* yang memadai untuk mengukur kemampuan generalisasi model. Tabel 4.9 menunjukkan pembagian data berdasarkan kategori sentimen.

**Tabel 4.9** Pembagian Data *Training* dan *Testing*

<b>Kelas</b>	<b>Jumlah</b>	<b>Data Training (80%)</b>	<b>Data Testing (20%)</b>
Positif	93	80	13
Negatif	892	708	184
	985	788	197

Data *training* digunakan untuk melatih model klasifikasi, sedangkan data *testing* digunakan untuk mengevaluasi akurasi model dalam mengklasifikasikan sentimen publik terhadap kebijakan kenaikan UKT. Dengan pembagian ini, diharapkan model dapat mengidentifikasi dan memahami pola sentimen dengan lebih akurat serta memberikan hasil klasifikasi yang dapat dijadikan referensi dalam analisis kebijakan pendidikan tinggi. Dalam penelitian ini digunakan *5-fold Cross Validation* ( $cv = 5$ ), artinya data *training* dibagi menjadi lima bagian, dan pelatihan model dilakukan sebanyak lima kali dengan pergiliran data *testing*. Teknik ini sangat bermanfaat dalam mencegah *overfitting*, karena model diuji pada data yang berbeda di setiap iterasinya. *Cross validation* juga memberikan gambaran performa model yang lebih akurat dibandingkan evaluasi menggunakan satu pembagian data saja.

## 4.6 Klasifikasi SVM

### 4.6.1 Klasifikasi SVM Manual terhadap Sampel Data

#### 1. Perhitungan TF-IDF

TF-IDF merupakan salah satu teknik dalam pemrosesan teks yang digunakan untuk menilai tingkat kepentingan suatu kata dalam sebuah dokumen. Pada bagian ini, dilakukan perhitungan manual TF-IDF berdasarkan sampel data yang disajikan pada Tabel 4.10.

**Tabel 4.10** Sampel Data *Tweet*

<b>D</b>	<b><i>Tweet</i></b>
D1	ukt naik sejahtera pendidik naik
D2	sedih ukt naik drastis
D3	ukt naik mahasiswa baru

Setelah sampel diperoleh, setiap dokumen diuraikan menjadi term atau kata. Setiap term dalam data akan dihitung untuk memperoleh bobot berdasarkan frekuensi kemunculan term dalam satu dokumen (TF) serta frekuensi kemunculan term dalam keseluruhan kumpulan dokumen (IDF), yang ditentukan menggunakan persamaan (2.2) dan persamaan (2.3), sebagai berikut:

$$tf(ukt, D_1) = \frac{f(t, D)}{\text{jumlah kata dalam dokumen } (D)} = \frac{1}{5} = 0,2$$

$$tf(naik, D_1) = \frac{f(t, D)}{\text{jumlah kata dalam dokumen } (D)} = \frac{2}{5} = 0,4$$

$$tf(ukt, D_2) = \frac{f(t, D)}{\text{jumlah kata dalam dokumen } (D)} = \frac{1}{4} = 0,25$$

$$IDF(ukt) = \log\left(\frac{N}{df(t)}\right) = \log\left(\frac{3}{3}\right) = 0$$

$$IDF(sejahtera) = \log\left(\frac{N}{df(t)}\right) = \log\left(\frac{1}{3}\right) = 0,477$$

Hasil perhitungan lengkap untuk sampel data dapat dilihat pada Tabel 4.11.

**Tabel 4.11** Hasil Perhitungan Frekuensi Kemunculan Term, TF dan IDF(t)

Kata	f(t,D)			TF			DF(t)	IDF(t)
	D1	D2	D3	D1	D2	D3		
ukt	1	1	1	0,2	0,25	0,25	3	0
naik	2	1	1	0,4	0,25	0,25	3	0
sejahtera	1	0	0	0,2	0	0	1	0,477
pendidik	1	0	0	0,2	0	0	1	0,477
sedih	0	1	0	0	0,25	0	1	0,477
drastis	0	1	0	0	0,25	0	1	0,477
mahasiswa	0	0	0	0	0	0,25	1	0,477
baru	0	0	0	0	0	0,25	1	0,477

Perhitungan berdasarkan persamaan (2.4) akan dilakukan untuk memperoleh nilai TF-IDF. Selanjutnya, dilakukan proses normalisasi menggunakan *Euclidean norm* sesuai dengan persamaan (2.5). Normalisasi ini bertujuan untuk menyamakan rentang nilai bobot yang diperoleh sehingga lebih terstruktur. Perhitungan TF-IDF dan normalisasi sebagai berikut :

$$V(\text{ukt}, D_1) = tf(t) \times idf(t) = 0,2 \times 0 = 0$$

$$V(\text{sejahtera}, D_1) = tf(t) \times idf(t) = 0,2 \times 0,477 = 0,095$$

$$V(\text{ukt}, D_2) = tf(t) \times idf(t) = 0,25 \times 0 = 0$$

$$V(\text{sedih}, D_2) = tf(t) \times idf(t) = 0,25 \times 0,477 = 0,119$$

$$\begin{aligned} V_{norm}(\text{sejahtera}, D_1) &= \frac{V_n}{\sqrt{V_1^2 + V_2^2 + \dots + V_n^2}} = \frac{0,095}{\sqrt{0^2 + 0^2 + 0,095^2 + \dots + 0^2}} \\ &= \frac{0,095}{\sqrt{0,00901 + 0,00901}} = \frac{0,095}{0,1349} = 0,706 \end{aligned}$$

Hasil akhir perhitungan lengkap untuk sampel data TF-IDF serta normalisasi vektor dapat dilihat pada Tabel 4.12.

**Tabel 4.12** Hasil Perhitungan TF-IDF

Kata	TF-IDF			Norm		
	D1	D2	D3	D1	D2	D3
ukt	0	0	0	0	0	0
naik	0	0	0	0	0	0
sejahtera	0,095	0	0	0,704	0	0
pendidik	0,095	0	0	0,704	0	0
sedih	0	0,119	0	0	0,706	0
drastis	0	0,119	0	0	0,706	0
mahasiswa	0	0	0,119	0	0	0,706
baru	0	0	0,119	0	0	0,706

## 2. *Imbalanced Data*

Pada penelitian ini, teknik SMOTE digunakan untuk menangani permasalahan ketidakseimbangan kelas dalam data *training*. Ketidakseimbangan kelas terjadi ketika jumlah data pada kelas positif dan negatif tidak seimbang, sehingga model cenderung mempelajari pola dari kelas mayoritas dan mengabaikan karakteristik kelas minoritas.

SMOTE bekerja dengan menghasilkan sampel sintetis dari kelas minoritas berdasarkan interpolasi nilai-nilai fitur dari data terdekat dalam ruang vektor. Teknik ini tidak hanya memperbanyak data secara acak, tetapi mempertahankan struktur karakteristik data aslinya sehingga dapat meningkatkan kemampuan model dalam mengenali pola dari kelas minoritas.

SMOTE hanya diterapkan pada data *training* dan tidak diterapkan pada data *testing*. Hal ini bertujuan agar evaluasi model tetap valid dan mencerminkan distribusi data yang sesungguhnya. Jika SMOTE juga dilakukan pada data *testing*, maka hasil evaluasi model dapat menjadi bias dan tidak mencerminkan performa model dalam kondisi yang sebenarnya.

Namun demikian, teknik SMOTE tidak diterapkan pada tahapan perhitungan manual. Pada bagian perhitungan manual, peneliti hanya menggunakan sampel data terbatas yang tidak melalui proses penyeimbangan kelas. Hal ini dimaksudkan untuk mempermudah pemaparan proses klasifikasi secara matematis, mulai dari perhitungan TF-IDF hingga penerapan *kernel linear* pada algoritma SVM. Oleh karena itu, data yang digunakan pada tahapan tersebut tetap mengikuti proporsi distribusi awal dan tidak mengalami perlakuan penyeimbangan.

### 3. Perhitungan SVM

Pada perhitungan manual, klasifikasi SVM dilakukan menggunakan *kernel linear* yang dirumuskan pada persamaan (2.17). Proses ini dimulai dari tahap pelatihan, dengan tujuan untuk menentukan *hyperplane* terbaik yang memisahkan data berdasarkan kelasnya. Dalam contoh ini, digunakan tiga data, yaitu dua data *training* dan satu data *testing*. Data  $x_1$  dan  $x_2$ , yang telah melalui proses normalisasi, digunakan sebagai data *training*. Sentimen positif dalam data *training* diberi label 1, sedangkan sentimen negatif diberi label -1. Data *training* yang digunakan ditampilkan pada Tabel 4.13.

**Tabel 4.13** Data *Training*

<i>term</i>	$x_1$	$x_2$
$term_1$	0	0
$term_2$	0	0
$term_3$	0,704	0
$term_4$	0,704	0
$term_5$	0	0,706
$term_6$	0	0,706
$term_7$	0	0
$term_8$	0	0
Y	1	-1

Berdasarkan data *training*, langkah selanjutnya adalah menghitung nilai matriks *kernel* K, untuk mengukur tingkat kemiripan antar data. Karena terdapat dua data *training*, maka matriks *kernel* yang dibentuk berukuran  $2 \times 2$  dengan elemen-elemen matriks dihitung menggunakan rumus *kernel linear* pada persamaan (2.17) sebagai berikut:

$$K(x_i, x_j) = x_i^T \cdot x_j$$

Dengan demikian, perhitungan elemen-elemen pada matriks *kernel* dilakukan sebagai berikut:

$$\begin{aligned} K(\mathbf{x}_1, \mathbf{x}_1) &= \mathbf{x}_1^T \cdot \mathbf{x}_1 = 0^2 + 0^2 + 0,704^2 + 0,704^2 + 0^2 + 0^2 + 0^2 + 0^2 \\ &= 0,9912 \end{aligned}$$

$$\begin{aligned} K(\mathbf{x}_1, \mathbf{x}_2) &= \mathbf{x}_1^T \cdot \mathbf{x}_2 = 0^2 + 0^2 + (0 \times 0,704)(0 \times 0,704) \\ &\quad + (0 \times 0,706) + (0 \times 0,706) + 0^2 + 0^2 = 0 \end{aligned}$$

$$K(\mathbf{x}_2, \mathbf{x}_1) = K(\mathbf{x}_1, \mathbf{x}_2) = 0$$

$$\begin{aligned} K(\mathbf{x}_2, \mathbf{x}_2) &= \mathbf{x}_2^T \cdot \mathbf{x}_2 = 0^2 + 0^2 + 0^2 + 0^2 + 0,706^2 + 0,706^2 + 0^2 + 0^2 \\ &= 0,9969 \end{aligned}$$

Sehingga, bentuk akhir dari matriks *kernel* K adalah sebagai berikut:

$$\mathbf{K}_{train} = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & K(\mathbf{x}_1, \mathbf{x}_2) \\ K(\mathbf{x}_2, \mathbf{x}_1) & K(\mathbf{x}_2, \mathbf{x}_2) \end{bmatrix} = \begin{bmatrix} 0,9912 & 0 \\ 0 & 0,9969 \end{bmatrix}$$

Selanjutnya, perhitungan dilakukan untuk memperoleh nilai *hyperplane* terbaik dengan menggunakan Persamaan *Langrange Multiplier* ( $\alpha$ ) yang dinyatakan dalam persamaan (2.12):

$$L = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

Dengan substitusi nilai yang telah diperoleh dari matriks *kernel* K, persamaan (2.12) menjadi:

$$\begin{aligned} L &= (\alpha_1 + \alpha_2) - \frac{1}{2} \{ (\alpha_1 \alpha_1 y_1 y_1 K(\mathbf{x}_1, \mathbf{x}_1)) + (\alpha_1 \alpha_2 y_1 y_2 K(\mathbf{x}_1, \mathbf{x}_2)) + \\ &\quad (\alpha_2 \alpha_1 y_2 y_1 K(\mathbf{x}_2, \mathbf{x}_1)) + (\alpha_2 \alpha_2 y_2 y_2 K(\mathbf{x}_2, \mathbf{x}_2)) \} \end{aligned}$$

$$0 = (\alpha_1 + \alpha_2) - \frac{1}{2}\{(\alpha_1\alpha_1(1) \times (0,9912)) + (\alpha_1\alpha_2(-1) \times (0)) + (\alpha_2\alpha_1(-1) \times (0)) + (\alpha_2\alpha_2(1) \times (0,9969))\}$$

Berdasarkan kondisi  $\sum_{i=1}^l \alpha_i y_i = 0$ , diperoleh:

$$\alpha_1(1) + \alpha_2(-1) = 0$$

$$\alpha_1 - \alpha_2 = 0$$

$$\alpha_1 = \alpha_2$$

Dengan demikian, persamaan (2.12) dapat disederhanakan menjadi:

$$L = (\alpha_1 + \alpha_1)$$

$$- \frac{1}{2}\{(\alpha_1\alpha_1(1 \times 0,9912)) + (\alpha_1\alpha_1(-1 \times 0))$$

$$+ (\alpha_1\alpha_1(-1 \times 0)) + (\alpha_1\alpha_1(1 \times 0,9969))\}$$

$$0 = (2\alpha_1) - \frac{1}{2}(0,9912\alpha_1\alpha_1 + 0,9969\alpha_1\alpha_1)$$

$$2\alpha_1 = \frac{1}{2}(1,9881\alpha_1\alpha_1)$$

$$2\alpha_1 = 0,99405\alpha_1\alpha_1$$

$$2 = 0,99405\alpha_1$$

$$\alpha_1 = 2,012 = \alpha_2$$

Setelah nilai  $\alpha_1 = \alpha_2$  diperoleh, langkah berikutnya adalah menghitung nilai  $\mathbf{w}$  menggunakan persamaan (2.15).

$$\mathbf{w} = \alpha_1 y_1 \mathbf{x}_1 + \alpha_2 y_2 \mathbf{x}_2$$

$$= \left( 2,012 \times 1 \times \begin{bmatrix} 0 \\ 0 \\ 0,704 \\ 0,704 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right) + \left( 2,012 \times (-1) \times \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0,706 \\ 0,706 \\ 0 \\ 0 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 0 \\ 0 \\ 1,416 \\ 1,416 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1,420 \\ -1,420 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1,416 \\ 1,416 \\ 0 \\ -1,420 \\ -1,420 \\ 0 \end{bmatrix}$$

Kemudian, nilai bias ( $b$ ) dihitung menggunakan persamaan (2.16):

$$b = \frac{1}{2} \left\{ \left( \frac{1}{y_1} - (\alpha_1 y_1 x_1^t x_1) + (\alpha_2 y_2 x_1^t x_2) \right) \right. \\ \left. + \left( \frac{1}{y_2} - (\alpha_1 y_1 x_2^t x_1) + (\alpha_2 y_2 x_2^t x_2) \right) \right\}$$

$$= \frac{1}{2} \left\{ \left( \frac{1}{-1} - (2,012 \times 1 \times 0,9912) + (2,012 \times (-1) \times 0) \right) \right. \\ \left. + \left( \frac{1}{1} - (2,012 \times 1 \times 0) + (2,012 \times (-1) \times 0,996872) \right) \right\}$$

$$= \frac{1}{2} (-1 - 1,99429 + 1 - 2,0057)$$

$$= \frac{1}{2} (-2,99429 + -1,0057)$$

$$= \frac{1}{2} (-4,0000008)$$

$$\approx -2$$

Setelah memperoleh nilai *Lagrange Multiplier*, *weight* dan bias dari proses *training*, tahap selanjutnya adalah melakukan klasifikasi dengan menggunakan data *testing*. Langkah ini bertujuan untuk mengidentifikasi apakah suatu *tweet* dalam data X terkait kenaikan UKT termasuk dalam kategori sentimen positif atau negatif. Sampel data *testing* yang digunakan adalah data  $x_{test}$  yang telah dinormalisasi, sebagaimana ditampilkan dalam Tabel 4.14.

**Tabel 4.14** Data *Testing*

<i>term</i>	$x_{test}$
$term_1$	0
$term_2$	0
$term_3$	0
$term_4$	0
$term_5$	0
$term_6$	0
$term_7$	0,706
$term_8$	0,706

Selanjutnya, menghitung nilai matriks *kernel* K menggunakan persamaan *linear* pada persamaan (2.17) pada data *testing* sebagai berikut:

$$K(x_1, x_{test}) = 0^2 + 0^2 + (0 \times 0,704) + (0 \times 0,704) + 0^2 + 0^2 \\ + (0 \times 0,706) + (0 \times 0,706) = 0$$

$$K(x_2, x_{test}) = 0^2 + 0^2 + 0^2 + 0^2 + (0 \times 0,706) + (0 \times 0,706) \\ + (0 \times 0,706) + (0 \times 0,706) = 0$$

Selanjutnya, dilakukan perhitungan untuk menentukan fungsi pemisah optimal dengan menggunakan persamaan (2.14).

$$\begin{aligned}
f(x_{test}) &= \text{sign}\left(\sum_{i=1}^l \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b\right) \\
&= \text{sign}(\alpha_1 y_1 K(\mathbf{x}_1, \mathbf{x}_{test}) + \alpha_2 y_2 K(\mathbf{x}_2, \mathbf{x}_{test}) + b) \\
&= \text{sign}(((2,012) \times (1) \times (0)) + ((2,012) \times (-1) \times (0)) + (-2)) \\
&= \text{sign}(0 + 0 - 2) \\
&= \text{sign}(-2) \\
&= -1
\end{aligned}$$

Berdasarkan hasil perhitungan pada sampel data *testing*, data tersebut diklasifikasikan sebagai kelas negatif karena nilai pemisah optimal yang diperoleh lebih kecil dari nol. Selanjutnya, proses perhitungan yang sama diterapkan pada seluruh data *training* untuk memperoleh model SVM yang optimal dalam klasifikasi sentimen.

#### 4.6.2 Klasifikasi SVM terhadap Seluruh Data

Klasifikasi terhadap data sentimen publik mengenai kenaikan UKT dilakukan dengan metode SVM yang diimplementasikan menggunakan bahasa pemrograman *Python* pada platform *Google Colab*. *Script* klasifikasi dapat dilihat pada Lampiran 4.

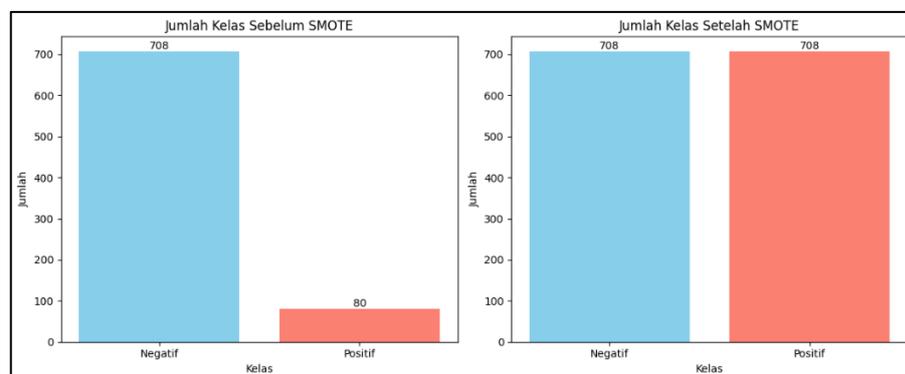
##### 1. Pelatihan Model

Data yang telah melalui tahap *preprocessing*, kemudian dibagi menjadi dua bagian, yaitu 80% sebagai data *training* dan 20% sebagai data *testing*. Data *training* digunakan untuk membangun dan melatih model klasifikasi, sedangkan data *testing* digunakan untuk mengukur kinerja model yang telah dibentuk. Setelah data dibagi, tahap selanjutnya adalah representasi data teks

ke dalam bentuk numerik menggunakan metode TF-IDF lalu dinormalisasikan. Hasil normalisasi TF-IDF kemudian menjadi *input* pada proses klasifikasi.

Sebelum model dilatih, dilakukan penyeimbangan distribusi kelas sentimen pada data *training*. Berdasarkan hasil pembagian data *training*, jumlah data kelas negatif sebanyak 708 sampel, sedangkan jumlah data kelas positif hanya 80 sampel. Ketidakseimbangan ini menunjukkan bahwa data bersifat *imbalanced* dan dapat menyebabkan bias dalam proses pembelajaran model. Untuk mengatasi hal tersebut, dilakukan penerapan SMOTE dengan tujuan meningkatkan jumlah sampel pada kelas minoritas melalui proses interpolasi terhadap nilai-nilai fitur berdasarkan tetangga terdekat dalam ruang vektor.

Setelah diterapkan metode SMOTE, jumlah data pada kelas positif meningkat menjadi 708 sampel, sehingga kedua kelas memiliki jumlah data yang sama. Dengan demikian, model dapat mempelajari pola sentimen dengan lebih baik dan tidak terpengaruh oleh ketimpangan jumlah data antar kelas. Perbandingan pembagian data *training* sebelum dan sesudah menerapkan SMOTE ditunjukkan pada Gambar 4.3.



**Gambar 4.3** Pembagian Data Sebelum dan Sesudah SMOTE

Proses pemodelan klasifikasi menggunakan SVM sangat dipengaruhi oleh pemilihan *kernel* dan *tuning* parameter. Pemilihan parameter terbaik menggunakan *grid search* dan *cross validation* untuk mengukur performa model. Setelah data seimbang, dilakukan pelatihan model SVM dengan empat jenis *kernel*, yaitu *linear*, polinomial, RBF dan *sigmoid*, dengan berbagai kombinasi parameter, yaitu nilai  $C$  dalam rentang  $10^{-1}$  hingga  $10^1$ ,  $\gamma$  dalam rentang  $10^{-3}$  hingga  $10^{-1}$ , dan  $\gamma$  dalam rentang 2 – 4, untuk memperoleh performa optimal. Setelah model SVM terbentuk dengan menggunakan data *training*, maka diperoleh nilai akurasi rata-rata dari *cross validation* untuk masing-masing kombinasi parameter yang disajikan pada tabel berikut:

**Tabel 4.15** Nilai Akurasi Parameter Model SVM *Kernel Linear*

<i>Kernel</i>	Parameter		Rata-rata Akurasi
	$C$		
<i>Linear</i>	10		0,98658
<i>Linear</i>	1		0,983763
<i>Linear</i>	0,1		0,863024

**Tabel 4.16** Nilai Akurasi Parameter Model SVM *Kernel RBF*

<i>Kernel</i>	Parameter		Rata-rata Akurasi
	$C$	$\gamma$	
RBF	10	0,1	0,9887
RBF	1	0,1	0,97108
RBF	0,1	0,1	0,55215
RBF	10	0,01	0,96683
RBF	1	0,01	0,54225
RBF	0,1	0,01	0,54225
RBF	10	0,001	0,61433
RBF	1	0,001	0,60515
RBF	0,1	0,001	0,54225

**Tabel 4.17** Nilai Akurasi Parameter Model SVM *Kernel Sigmoid*

Kernel	Parameter		Rata-rata Akurasi
	$C$	$\gamma$	
<i>Sigmoid</i>	10	0,1	0,98376
<i>Sigmoid</i>	1	0,1	0,86302
<i>Sigmoid</i>	0,1	0,1	0,54225
<i>Sigmoid</i>	10	0,01	0,86302
<i>Sigmoid</i>	1	0,01	0,54225
<i>Sigmoid</i>	0,1	0,01	0,54225
<i>Sigmoid</i>	10	0,001	0,54225
<i>Sigmoid</i>	1	0,001	0,54225
<i>Sigmoid</i>	0,1	0,001	0,54225

**Tabel 4.18** Nilai Akurasi Parameter Model SVM *Kernel Polinomial*

<i>Kernel</i>	Parameter			Rata-rata Akurasi	<i>Kernel</i>	Parameter			Rata-rata Akurasi
	$C$	$\gamma$	$d$			$C$	$\gamma$	$d$	
Poli	10	0,1	2	0,71895	Poli	1	0,01	3	0,55633
Poli	0,1	0,001	4	0,57535	Poli	10	0,1	3	0,55633
Poli	0,1	0,01	4	0,57464	Poli	0,1	0,001	3	0,55633
Poli	1	0,1	4	0,57464	Poli	0,1	0,01	3	0,55633
Poli	10	0,1	4	0,57464	Poli	1	0,001	3	0,55633
Poli	10	0,01	4	0,57464	Poli	1	0,01	2	0,53239
Poli	10	0,001	4	0,57464	Poli	1	0,001	2	0,53239
Poli	1	0,01	4	0,57464	Poli	0,1	0,01	2	0,53239
Poli	1	0,001	4	0,57464	Poli	0,1	0,1	2	0,53239
Poli	0,1	0,1	4	0,57464	Poli	0,1	0,001	2	0,53239
Poli	0,1	0,1	3	0,55633	Poli	1	0,1	2	0,53239
Poli	10	0,001	3	0,55633	Poli	10	0,001	2	0,53239
Poli	10	0,01	3	0,55633	Poli	10	0,01	2	0,53239
Poli	1	0,1	3	0,55633					

Dari hasil pelatihan model SVM, dapat disimpulkan bahwa kombinasi parameter terbaik untuk masing-masing *kernel* adalah sebagai berikut:

**Tabel 4.19** Perbandingan Performa Setiap *Kernel*

<i>Kernel</i>	Parameter			Rata-rata Akurasi
	<i>C</i>	$\gamma$	<i>d</i>	
<i>Linear</i>	10	-	-	0,98658
<b>RBF</b>	<b>10</b>	<b>0,1</b>	-	<b>0,9887</b>
Poli	10	0,1	2	0,71895
<i>Sigmoid</i>	10	0,1	-	0,98376

Berdasarkan Tabel 4.19, *kernel* RBF dengan parameter  $C = 10$  dan  $\gamma = 0,1$  menghasilkan performa terbaik dengan rata-rata akurasi sebesar 0,9887. Oleh karena itu, model SVM terbaik yang dipilih adalah model dengan *kernel* RBF, menggunakan parameter  $C = 10$  dan  $\gamma = 0,1$ .

## 2. Evaluasi Model

Setelah model terbaik diperoleh dari hasil pelatihan, model diuji pada data *testing* yang sebelumnya tidak disertakan dalam proses pelatihan. Data *testing* juga melalui tahap TF-IDF, namun tidak dikenai proses SMOTE, agar uji performa lebih mencerminkan kondisi data aktual. Evaluasi dilakukan untuk mengetahui seberapa baik model mampu menggeneralisasi pola yang telah dipelajari. Berikut merupakan hasil *confusion matrix* yang didapatkan dari model klasifikasi menggunakan SVM pada Tabel 4.20.

**Tabel 4.20** *Confusion Matrix*

Kelas		Kelas Aktual	
		Positif	Negatif
Kelas Prediksi	Positif	2	2
	Negatif	11	182

Berdasarkan hasil pengujian model klasifikasi SVM terhadap data sentimen publik terkait kebijakan kenaikan UKT, diperoleh *confusion matrix* yang

menunjukkan bahwa dari total 197 data *testing*, sebanyak 2 data diklasifikasikan dengan benar sebagai kelas positif (TP), dan 182 data diklasifikasikan dengan benar sebagai kelas negatif (TN). Namun, terdapat 2 data yang seharusnya termasuk dalam kelas negatif tetapi diprediksi sebagai kelas positif (FP), serta 11 data yang seharusnya termasuk dalam kelas positif namun diprediksi sebagai kelas negatif (FN). Hasil prediksi di atas dapat digunakan untuk mendapatkan nilai akurasi pada persamaan (2.19), presisi pada persamaan (2.20), *recall* pada persamaan (2.21), *F1-score* pada persamaan (2.22) dan ROC AUC pada persamaan (2.23), sehingga didapatkan hasil sebagai berikut:

a. Nilai Akurasi

$$Akurasi = \frac{2 + 182}{2 + 2 + 182 + 11} \times 100\% = \frac{184}{197} \approx 93\%$$

Nilai akurasi yang diperoleh sebesar 93% menunjukkan bahwa model mampu melakukan klasifikasi dengan benar terhadap 93% data, yang menunjukkan bahwa model dapat diandalkan dalam melakukan klasifikasi secara umum.

b. Nilai Presisi

$$Presisi = \frac{2}{2 + 2} \times 100\% = \frac{2}{4} = 50\%$$

Nilai presisi yang diperoleh sebesar 50% menunjukkan bahwa hanya setengah dari data yang diprediksi sebagai sentimen positif oleh model benar-benar merupakan sentimen positif.

c. Nilai *Recall*

$$Recall = \frac{2}{2 + 11} \times 100\% \approx 15,4\%$$

Nilai *recall* yang rendah, yaitu 15,4%, yang berarti hanya sebagian kecil data sentimen positif yang berhasil dikenali oleh model. Rendahnya nilai *recall* ini mengindikasikan bahwa model memiliki kesulitan dalam mendeteksi opini publik yang bersifat positif terhadap kebijakan kenaikan UKT. Hal ini disebabkan oleh tingginya jumlah data positif yang tidak berhasil terklasifikasi dengan benar (FN), sehingga berdampak pada rendahnya representasi opini positif dalam hasil analisis

d. Nilai *F1-score*

$$F1 - score = 2 \times \frac{50 \times 15,4}{50 + 15,4} \times 100\% \approx 23,5\%$$

Nilai *F1-score* yang diperoleh sebesar 23,5% menunjukkan bahwa model kurang mampu mengelola keseimbangan tersebut dengan baik. Nilai ini mencerminkan kinerja yang rendah dalam klasifikasi kelas positif, yang merupakan dampak dari presisi yang sedang dan *recall* yang rendah.

## e. Grafik ROC AUC

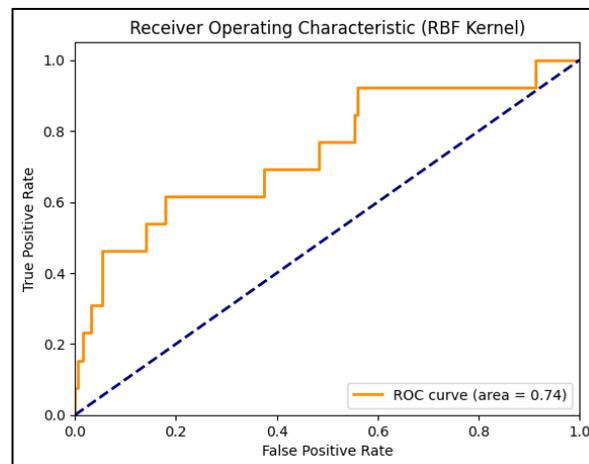
Selain menggunakan metrik akurasi, presisi, *recall*, dan *F1-Score*, evaluasi model juga dilakukan melalui kurva ROC dan nilai AUC. Dalam konteks ketidakseimbangan kelas seperti pada penelitian ini, nilai AUC menjadi metrik penting karena mampu mengukur performa klasifikasi secara menyeluruh tanpa bias terhadap kelas mayoritas. Nilai TPR dihitung menggunakan persamaan (2.23) dan FPR dihitung menggunakan persamaan

(2.24). Hasil perhitungan nilai FPR dan TPR untuk berbagai *threshold* klasifikasi ditampilkan pada Tabel 4.21, dengan nilai FPR dan TPR secara lengkap disajikan pada Lampiran 9.

**Tabel 4.21** Nilai FPR dan TPR

No	FPR	TPR
1	0	0
2	0	0,07692
3	0,00543	0,07692
4	0,00543	0,15385
...	...	...
26	0,91304	1
27	1	1

Setelah memperoleh nilai FPR dan TPR maka dilakukan pemetaan ke dalam grafik untuk membentuk kurva ROC pada gambar 4.4.



**Gambar 4.4** Kurva ROC dan Nilai AUC Model SVM

Setelah grafik ROC terbentuk melalui pemetaan nilai FPR dan TPR pada setiap *threshold*, langkah selanjutnya adalah menghitung AUC. Nilai AUC mencerminkan kemampuan model dalam membedakan antara kelas positif

dan negatif secara keseluruhan. Perhitungan nilai AUC menggunakan persamaan (2.25) sebagai berikut:

$$\begin{aligned}
 AUC &= \frac{1}{2} \sum_{i=1}^l (FPR_{i+1} + FPR_i)(TPR_{i+1} - TPR_i) \\
 &= \frac{1}{2} (((TPR_2 + TPR_1)(FPR_2 - FPR_1)) + ((TPR_3 + TPR_2)(FPR_3 \\
 &\quad - FPR_2)) + \dots + ((TPR_{29} + TPR_{28})(FPR_{29} - FPR_{28}))) \\
 &= \frac{1}{2} (((0,0769 + 0)(0 - 0)) + ((0,0769 + 0,0769)(0,0054 - 0)) + \dots \\
 &\quad + ((1 + 1)(1 - 0,9130))) \\
 &= \frac{1}{2} (0 + 0,00041 + \dots + 0,174) \\
 &\approx 0,74
 \end{aligned}$$

Berdasarkan kurva ROC pada gambar 4.4 dan nilai AUC yang diperoleh, terlihat bahwa kurva ROC berada di atas garis diagonal acuan. Hal ini menunjukkan bahwa model memiliki kemampuan yang lebih baik dari tebakan acak dalam mengklasifikasikan data. Luas area di bawah kurva ROC (AUC) sebesar 0,74 menunjukkan bahwa dalam 74% kasus, model mampu membedakan secara benar antara satu data dari kelas positif dan satu data dari kelas negatif yang dipilih secara acak. Perolehan nilai AUC sebesar 0,74, menunjukkan bahwa model SVM yang digunakan dalam penelitian ini termasuk dalam kategori cukup dalam membedakan antara sentimen positif dan negatif berdasarkan Tabel 2.6.

Berdasarkan seluruh metrik evaluasi, dapat disimpulkan bahwa model klasifikasi SVM yang digunakan dalam penelitian ini memiliki performa yang cukup baik secara umum, dengan nilai akurasi sebesar 93% dan nilai AUC

sebesar 0,74 yang menunjukkan kemampuan diskriminatif yang memadai. Meskipun masih terdapat kekurangan dalam mendeteksi kelas minoritas (sentimen positif), model ini tetap dapat dijadikan dasar yang layak dalam melakukan analisis sentimen publik terhadap kebijakan kenaikan UKT.

Penerapan klasifikasi sentimen publik terhadap kebijakan kenaikan UKT menggunakan algoritma SVM memberikan kontribusi yang signifikan dalam hal efisiensi waktu pengolahan data berskala besar (*big data*) serta peningkatan kualitas evaluasi kebijakan oleh pemerintah. Di era digital, opini dan tanggapan masyarakat terhadap suatu kebijakan banyak diungkapkan melalui media sosial dalam jumlah yang sangat besar dan juga tidak terstruktur. Pengolahan data secara manual akan memerlukan waktu yang lama serta sumber daya manusia yang besar. Oleh karena itu, penggunaan algoritma klasifikasi seperti SVM menjadi solusi karena mampu mengolah data dalam jumlah besar secara otomatis, efisien dan akurasi yang baik. Melalui proses transformasi data teks menjadi data numerik dengan TF-IDF, serta penerapan klasifikasi menggunakan model SVM, analisis sentimen dapat dilakukan dalam waktu yang relatif singkat dengan hasil yang akurat.

Kecepatan dan efisiensi sangat penting, terutama untuk pemerintah yang membutuhkan respons cepat terhadap dinamika opini publik yang berkembang secara *real-time*. Selain itu, hasil klasifikasi sentimen juga berperan dalam proses evaluasi kebijakan. Dengan mengetahui persepsi masyarakat secara objektif berdasarkan data yang telah dianalisis, pemerintah dapat menilai sejauh mana suatu kebijakan diterima atau ditolak oleh publik.

Dengan demikian, penerapan klasifikasi sentimen menggunakan algoritma SVM tidak hanya meningkatkan efisiensi waktu dalam pengolahan data, tetapi juga memberikan dukungan informasi yang akurat dan berbasis data dalam proses evaluasi kebijakan. Hasil analisis ini dapat menjadi acuan bagi pemerintah dalam merumuskan, merevisi, maupun menyempurnakan kebijakan secara lebih responsif terhadap aspirasi serta kebutuhan masyarakat.

#### **4.7 Analisis Sentimen terhadap Kebijakan Kenaikan Uang Kuliah Tunggal dalam Pandangan Islam**

Hasil pengujian model SVM terhadap data sentimen publik terkait kebijakan kenaikan UKT menunjukkan bahwa mayoritas sentimen masyarakat bersifat negatif. Hal ini tercermin dari jumlah data yang diklasifikasikan sebagai sentimen negatif yang jauh lebih tinggi dibandingkan sentimen positif, yakni sebanyak 182 dari 197 data berhasil diklasifikasikan dengan benar sebagai sentimen negatif, sedangkan hanya 2 data yang diklasifikasikan benar sebagai sentimen positif. Dominasi sentimen negatif mengindikasikan ketidakpuasan dan kekhawatiran masyarakat terhadap dampak kebijakan tersebut.

Dalam perspektif Islam, pendapat masyarakat memiliki nilai penting dalam proses pengambilan keputusan, terutama yang menyangkut kepentingan bersama seperti akses terhadap pendidikan. Islam mengajarkan prinsip musyawarah sebagai bentuk partisipasi dalam menentukan arah kebijakan. Sebagaimana ditegaskan dalam QS. Asy-Syura ayat 38.

Ayat ini menunjukkan bahwa Islam menghargai musyawarah dalam pengambilan keputusan. Sehingga, ketika mayoritas masyarakat menunjukkan

penolakan terhadap kebijakan, seperti dalam dominasi sentimen negatif terhadap kebijakan UKT, hal tersebut dapat dijadikan dasar evaluasi oleh para pengambil kebijakan. Respons masyarakat yang luas menunjukkan adanya kebutuhan untuk meninjau kembali dampak kebijakan terhadap keadilan sosial dan kesejahteraan umum.

Dalam Islam, menyampaikan pendapat juga harus disertai dengan adab yang baik. Kebebasan berpendapat bukan berarti bebas mencela, menghina, atau berkata kasar, baik secara lisan maupun dalam bentuk tulisan seperti komentar di media sosial. Sebagaimana juga ditegaskan dalam QS. Al-Hujurat 11. Komentar di media sosial termasuk dalam bentuk ekspresi lisan yang harus dijaga. Umat Islam dituntut untuk menjaga lisannya dan cara menyampaikan pendapat di ruang publik daring.

Dengan mempertimbangkan ayat-ayat Al-Qur'an dan pandangan para ulama, serta hasil analisis sentimen publik, dapat disimpulkan bahwa kebijakan kenaikan UKT perlu ditinjau kembali. Hal ini bertujuan agar kebijakan tersebut selaras dengan nilai-nilai Islam yang menekankan pentingnya musyawarah untuk kepentingan bersama. Pendidikan tinggi idealnya menjadi hak yang dapat diakses oleh seluruh lapisan masyarakat, bukan menjadi beban yang mempersempit kesempatan masyarakat untuk belajar.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Berdasarkan rumusan masalah dan pembahasan pada bab sebelumnya maka diperoleh kesimpulan sebagai berikut:

1. Klasifikasi sentimen publik terhadap kebijakan kenaikan Uang Kuliah Tunggal (UKT) di Platform X telah berhasil dilakukan menggunakan metode *Support Vector Machine* (SVM). Hasil klasifikasi menunjukkan bahwa dari 197 data *testing*, sebanyak 182 data berhasil diklasifikasikan dengan benar sebagai sentimen negatif, dan sebanyak 2 data yang diklasifikasikan benar sebagai sentimen positif. Namun demikian, terdapat 2 data yang seharusnya merupakan sentimen negatif namun diklasifikasikan sebagai positif, serta 11 data yang merupakan sentimen positif namun diklasifikasikan sebagai negatif.
2. Tingkat akurasi model klasifikasi SVM dalam mengklasifikasikan sentimen publik terhadap kebijakan kenaikan UKT mencapai 93%, yang menunjukkan bahwa secara keseluruhan model memiliki performa yang baik. Selain itu, nilai ROC AUC sebesar 74% menunjukkan kemampuan model yang cukup baik dalam membedakan antara kelas positif dan kelas negatif. Meskipun demikian, performa model dalam mendeteksi sentimen positif masih perlu ditingkatkan, mengingat ketidakseimbangan distribusi kelas pada data yang digunakan.

## 5.2 Saran

Adapun beberapa hal yang dapat dijadikan acuan untuk penyempurnaan penelitian selanjutnya maka penulis memberikan saran berikut:

1. Penanganan ketidakseimbangan data dapat dilakukan dengan menggunakan kombinasi metode yang lain, seperti metode *under-sampling*, *over-sampling* maupun kombinasi metode *over* dan *under sampling* agar model lebih baik.
2. Dalam pengembangan penelitian selanjutnya, disarankan untuk melakukan pengujian menggunakan kombinasi algoritma SVM. Salah satunya adalah SVM *Multi-Class* dengan menambahkan kategori kelas netral.
3. Disarankan untuk menambah jumlah data *training*, terutama data dengan label sentimen positif, agar distribusi kelas menjadi lebih seimbang.
4. Disarankan kepada pemerintah untuk mempertimbangkan kembali kebijakan kenaikan UKT dengan memperhatikan aspirasi publik. Hasil analisis sentimen dapat dimanfaatkan sebagai masukan dalam evaluasi kebijakan agar lebih responsif dan sesuai dengan kondisi sosial ekonomi masyarakat.

## DAFTAR PUSTAKA

- Al Azies, H., Trishnanti, D., & Mustikawati, E. (2019). Comparison of Kernel Support Vector Machine (SVM) in Classification of Human Development Index (HDI). *IPTEK Journal of Proceedings Series*, 6, 55–57.
- Ariyanto, R. A., & Chamidah, N. (2021). Sentiment Analysis for Zoning System Admission Policy Using Support Vector Machine and Naive Bayes Methods. *Journal of Physics: Conference Series*, 1776(1). <https://doi.org/10.1088/1742-6596/1776/1/012058>
- Astuti, S. P. (2020). *Analisis Sentimen Berbasis Aspek pada Aplikasi TOKOPEDIA Menggunakan LDA dan Naive Bayes*. UIN Syarif Hidayatullah Jakarta.
- Bahri, S., Bahri, P., & Lal, S. (2018). A Novel approach of Sentiment Classification using Emoticons. *Procedia Computer Science*, 132, 669–678. <https://doi.org/10.1016/j.procs.2018.05.067>
- Baita, A., Pristyanto, Y., & Cahyono, N. (2021). Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN). *Information System Journal (INFOS)*, 4(2), 43–44.
- BPS. (2023). *Jumlah Perguruan Tinggi1, Dosen, dan Mahasiswa2 (Negeri dan Swasta) di Bawah Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Menurut Provinsi, 2022*. [Www.Bps.Go.Id](http://www.bps.go.id). <https://www.bps.go.id/id/statistics-table/3/Y21kVGRHNXZVMEI3S3pCRllYMHJRbnB1WkVZemR6MDkjMw==/jumlah-perguruan-tinggi--tenaga-pendidik-dan-mahasiswa-negeri-dan-swasta--di-bawah-kementerian-ri-set--teknologi-dan-pendidikan-tinggi-kementerian-pendidikan-dan-kebudayaan-menurut-provinsi--2022.html?year=2022>
- Budiman, I., Faisal, M. R., & Nugrahadi, D. T. (2020). Studi Ekstraksi Fitur Berbasis Vektor Word2Vec pada Pembentukan Fitur Berdimensi Rendah. *Jurnal Komputasi*, 8(1), 62–68.
- Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Eko, P., Utomo, P., Khaira, U., & Suratno, T. (2019). Analisis Sentimen Online Review Pengguna Bukalapak Menggunakan Metode Algoritma TF-IDF. In *JUSS) Jurnal Sains dan Sistem Informasi* (Vol. 2, Issue 2).

- Ghoffar, M. A., Mu'thi, A., & Al-Atsari, A. I. (2004). *Tafsir Ibnu Katsir* (1st ed.). Pustaka Imam asySyafi'i . [www.pustakaimamsyafii.com](http://www.pustakaimamsyafii.com)
- Gifari, O. I., Adha, M., Rifky Hendrawan, I., Freddy, F., & Durrand, S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY)*, 2(1).
- Gu, S., Wang, F., Patel, N. P., Bourgeois, J. A., & Huang, J. H. (2019). A model for basic emotions using observations of behavior in *Drosophila*. In *Frontiers in Psychology* (Vol. 10, Issue APR). Frontiers Media S.A. <https://doi.org/10.3389/fpsyg.2019.00781>
- Han, J., Kamber, M., & Pei, J. (2022). *Data Mining Concepts and Techniques*.
- Herlinawati, N., Yuliani, Y., Faizah, S., Gata, W., Komputer STMIK Nusa Mandiri Jl Damai No, I., Jati Barat, W., & Selatan, J. (2020). Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine (Vol. 5, Issue 2).
- Iriananda, S. W., Budiawan, R. W., Rahman, A. Y., & Istiadi, I. (2024). Optimasi Klasifikasi Sentimen Komentar Pengguna Game Bergerak Menggunakan Svm, Grid Search Dan Kombinasi N-Gram. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(4), 743–752. <https://doi.org/10.25126/jtiik.1148244>
- Juwiantho, H., Setiawan, E. I., Santoso, J., Purnomo, M. H., Informasi, D. T., Tinggi, S., & Surabaya, T. (2020). Sentiment Analysis Twitter Bahasa Indonesia Berbasis Word2vec Menggunakan Deep Convolutional Neural Network. *J. Teknol. Inf. Dan Ilmu Komput.*, 7(1), 181–188. <https://doi.org/10.25126/jtiik.202071758>
- Khalid, R., & Javaid, N. (2020). A Survey on Hyperparameters Optimization Algorithms of Forecasting Models in Smart Grid. *Sustainable Cities and Society*, 6. <https://doi.org/https://doi.org/10.1016/j.scs.2020.102275>
- Kementerian Agama RI. 2024. *Quran Kemenag*.
- Mawardi, I. (2024, May 27). *Pernyataan Lengkap Nadiem Makarim Batalkan Kenaikan UKT Tahun Ini*. [Www.News.Detik.Com](http://www.News.Detik.Com).
- Nugroho, A. S. (2008). Support Vector Machine: Paradigma Baru dalam Softcomputing. In *Konferensi Nasional Sistem dan Informatika*.
- Nurhafida, S. I., & Sembiring, F. (2022). Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM). In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 6, Issue 1).

- Ramadhan, N. G., & Ramadhan, T. I. (2022). Analysis Sentiment Based on IMDB Aspects from Movie Reviews using SVM. *Jurnal Dan Penelitian Teknik Informatika*, 6(1), 39–45. <https://doi.org/10.33395/sinkron.v7i1.11204>
- Rangga, M., Nasution, A., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *JURNAL INFORMATIKA*, 6(2), 212–218. <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- Riefky, M., & Anandyani, A. R. (2019). Klasifikasi Persepsi Pengguna Twitter Terhadap Tuntutan Keringanan Pembayaran UKT pada Masa Pandemi COVID-19 Menggunakan K-nearest Neighbor.
- Roifa, A. N. (2018). *Text Mining Method for Classifying Naïve Bayes Classifier Based on the Content Of News*. Institut Teknologi of Sepuluh Nopember.
- Sarkar, D. (2016). Text Analytics with Python. In *Text Analytics with Python*. Apress. <https://doi.org/10.1007/978-1-4842-2388-8>
- Suyoga, I. G. S., Kencana, I. P. E. N., & Sukarsa, I. K. G. (2017). Penggolongan Uang Kuliah Tunggal Menggunakan *Support Vector Machine*. *E-Jurnal Matematika*, 6(4), 220. <https://doi.org/10.24843/mtk.2017.v06.i04.p169>
- Tando, A. G., & Irawan, M. I. (2023). Analisis Dinamika Harga Saham yang Dipengaruhi oleh Analisis Sentimen di Media Sosial Menggunakan Algoritma Support Vector Machine. *JURNAL SAINS DAN SENI ITS*, 12(1), A48.
- Vapnik, V., Cortes, C., & Saitta, L. (1995). *Support-Vector Networks Editor*. In *Machine Learning* (Vol. 20). Kluwer Academic Publishers.
- Wahyudi, R., Kusumawardhana, G., Purwokerto, A., Letjend, J., Soemarto, P., Purwanegara, K., Purwokerto, T., & Banyumas, K. (2021). Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *JURNAL INFORMATIKA*, 8(2). <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- Wahyuni, N. A., Ayu, D. P., & Irsyad, H. (2024). Analisis Sentimen di Youtube Terhadap Kenaikan UKT Menggunakan Metode Support Vector Machine. *Arcitech: Journal of Computer Science and Artificial Intelligence*, 4(1), 57. <https://doi.org/10.29240/arcitech.v4i1.10829>
- Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). *Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes*. <http://www.biomedcentral.com/1472-6947/10/16>

## LAMPIRAN

### Lampiran 1. *Script Crawling Data X*

```
# Import required Python package
!pip install pandas

# Install Node.js (because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --
dearmor -o /etc/apt/keyrings/nodesource.gpg
!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg]
https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro main" | sudo tee
/etc/apt/sources.list.d/nodesource.list
!sudo apt-get update
!sudo apt-get install nodejs -y
!node -v
# Crawl Data

filename = 'Data Tweet.csv'
search_keyword = 'ukt lang:id until:2024-02-25 since:2024-01-19 -filter:links -filter:replies'
limit = 1000

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l
{limit} --token {twitter_auth_token}
import pandas as pd

file_path = f'tweets-data/{filename}'
df = pd.read_csv(file_path, delimiter=",")
display(df)
```

### Lampiran 2. *Script Preprocessing text*

```
!pip install Sastrawi
!pip install -U textblob
!python -m textblob.download_corpora
import re, string # Untuk manipulasi teks dasar (regex dan karakter)
import nltk # Natural Language Toolkit untuk pemrosesan bahasa alami
from nltk.tokenize import word_tokenize # Tokenisasi kata
from nltk.corpus import stopwords # Daftar kata
import pandas as pd
import numpy as np # Operasi numerik
import matplotlib.pyplot as plt # Visualisasi grafik
import seaborn as sns # Visualisasi statistik
from wordcloud import WordCloud # Visualisasi word cloud
```

```

from textblob import TextBlob
from transformers import pipeline
from transformers import AutoTokenizer, AutoModelForSequenceClassification
from sklearn.model_selection import train_test_split # Membagi data latih dan uji
from sklearn.feature_extraction.text import TfidfVectorizer # Mengubah teks ke vektor TF-IDF
from sklearn.preprocessing import LabelEncoder # Mengubah label kategori ke numerik
from sklearn.preprocessing import Normalizer # Normalisasi data
from imblearn.over_sampling import SMOTE # Oversampling kelas minoritas
from sklearn.svm import SVC # Support Vector Classifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, roc_auc_score #evaluasi model
from sklearn.model_selection import GridSearchCV # Hyperparameter tuning
nltk.download('stopwords')
nltk.download('punkt')

# memasukkan data tweet
def load_data():
    data = pd.read_csv('/content/Data tweet (3).csv')
    return data
tweet_df = load_data()

# memasukkan label dari setiap data tweet
labelling_df = pd.read_excel('/content/Label.xlsx')
tweet_df = pd.merge(tweet_df, labelling_df[['Full_text', 'Label']])

# tahapan preprocessing
#1 casefolding
def lowercase(text):
    low = text.lower()
    return low
tweet_df['lower_text'] = tweet_df['Full_text'].apply(lambda low:lowercase(str(low)))

#2 cleaning
def preprocess(text):
    text = text.strip() # Menghilangkan spasi di awal dan akhir
    emoji_pattern = re.compile("[\u0001F600-\u0001F64F\u0001F300-\u0001F5FF\u0001F680-\u0001F6FF\u0001F1E0-\u0001F1FF\u00002500-\u00002BEF\u00002702-\u000027B0\u00002702-\u000027B0\u000024C2-\u0001F251\u0001f926-\u0001f937\u00010000-\u0010ffff\u00002640-\u00002642\u00002600-\u00002B55\u0000200d\u000023cf\u000023e9\u0000231a\u0000fe0f\u00003030]#", flags=re.UNICODE)
    text = re.sub(emoji_pattern, "", text)
    text = emoji_pattern.sub("", text)
    text = re.sub(r'<.*?>', "", text) # Menghilangkan HTML tags

```

```

text = re.sub(r'(\w+)\^\S+|\1_\S+', r'\1', text) # Menghilangkan karakter '@' beserta kata
yang mengikutinya
text = re.sub(r'@\S+\s*', "", text) # Menghilangkan URL atau tautan
text = re.sub(r'https?:/\S+', "", text)
text = re.sub(r'%s]' % re.escape(string.punctuation), '', text) # Menghilangkan tanda baca
text = re.sub(r'\s+', ' ', text) # Mengganti beberapa spasi dengan satu spasi
text = re.sub(r'\d', '', text) # Menghilangkan angka
text = re.sub(r'\[[0-9]*\]', '', text) # Menghilangkan referensi dalam tanda kurung
text = re.sub(r'\s+', ' ', text) # Mengganti beberapa spasi dengan satu spasi lagi
return text
tweet_df['cleaning_text'] = tweet_df['lower_text'].apply(lambda pre: preprocess(str(pre)))

# Remove duplicate rows based on the 'Full_text' column
tweet_df.drop_duplicates(subset=['Full_text'], inplace=True)

#3 Normalisasi
# Memuat kamus slang dari file Excel
slang_df = pd.read_excel('/content/Slank.xlsx')

# Membuat kamus slang dari DataFrame, memastikan nilai-nilai diubah menjadi string
dict_slangs = dict(zip(slang_df['Kata Slank'], slang_df['Kata Baku'].astype(str)))

# Fungsi replace_slang tetap sama
def replace_slang(text, dict_slangs):
    words = TextBlob(text).words
    replaced_words = [dict_slangs.get(word, word) for word in words]
    return ''.join(replaced_words)

# Membersihkan teks dalam DataFrame dari kata slang
clean_text = []
for review in tweet_df['cleaning_text']:
    wordlist = TextBlob(review).words
    for k, v in enumerate(wordlist):
        if v in dict_slangs.keys():
            wordlist[k] = dict_slangs[v]
    clean_text.append(''.join(wordlist))

tweet_df['slangs_text'] = clean_text

#4 Tokenizing
def word_token(text):
    return word_tokenize(text)
tweet_df['token_text'] = tweet_df['slangs_text'].apply(lambda
tokenize:word_tokenize(str(tokenize)))

#5 Stopword Removal
indo_stopwords = pd.read_excel('/content/Stopwords.xlsx')

```

```

indo_stopwords = indo_stopwords['stopwords_indonesia'].tolist() # Ubah ke list

# Fungsi untuk menghapus stopwords dari teks
def remove_stopwords(text, indo_stopwords):
    filtered_text = []
    for token in text:
        if token not in indo_stopwords:
            filtered_text.append(token)
    return filtered_text

# Menerapkan fungsi remove_stopwords ke kolom 'token_text'
tweet_df['stopword_text'] = tweet_df['token_text'].apply(lambda stop:
remove_stopwords(stop, indo_stopwords)) # Gunakan indonesian_stopwords

#6 Stemming
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

def stem_text(text):
    stemmer = StemmerFactory().create_stemmer()
    return [stemmer.stem(token) for token in text]

# Menerapkan fungsi stem_text ke kolom 'stopword_text'
tweet_df['stemmed_text'] = tweet_df['stopword_text'].apply(stem_text)

# Fungsi untuk melakukan rejoin pada teks yang sudah distem
def rejoin_text(tokens):
    return ' '.join(tokens)

# Melakukan rejoin pada kolom 'stemmed_text'
tweet_df['rejoined_text'] = tweet_df['stemmed_text'].apply(rejoin_text)

# Simpan hasil dengan nama 'tweet_data.csv'
tweet_df.to_csv('tweet_data.csv', index=False)

```

### Lampiran 3. *Script* Visualisasi Data

```

# Menggabungkan semua kata dalam satu list
all_words = [word for sublist in tweet_df['stopword_text'] for word in sublist]

# Menggabungkan semua kata menjadi satu teks
all_text = ' '.join(all_words)

# Membuat objek WordCloud
wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(all_text)

# Menampilkan WordCloud

```

```
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

#### Lampiran 4. Script Klasifikasi SVM

```
# 1. Pembagian Data
X = tweet_df['rejoined_text']
y = tweet_df['Label']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 2. TF-IDF pada data training dan testing
tfidf_vectorizer = TfidfVectorizer()
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_test_tfidf = tfidf_vectorizer.transform(X_test)

#3. Melakukan Normalisasi menggunakan Euclidean Normalization pada hasil TF-IDF
normalizer = Normalizer(norm='l2')
X_train_tfidf = normalizer.fit_transform(X_train_tfidf)
X_test_tfidf = normalizer.transform(X_test_tfidf)

# 3. Mengganti label dengan kode angka
label_encoder = LabelEncoder()
y_train = label_encoder.fit_transform(y_train)
y_test = label_encoder.transform(y_test)

# 4. Melakukan SMOTE pada data training
smote = SMOTE(random_state=42)
X_train_resampled, y_train_resampled = smote.fit_resample(X_train_tfidf, y_train)

# Distribusi data
print("Before SMOTE:")
print(pd.Series(y_train).value_counts()) # Distribusi data sebelum SMOTE
print("\nAfter SMOTE:")
print(pd.Series(y_train_resampled).value_counts()) # Distribusi data setelah SMOTE

# 5. Melakukan klasifikasi menggunakan SVM Train the SVM Classifier
# Define the parameter grid for each kernel
param_grid = {
    'linear': {'kernel': ['linear'], 'C': [0.1, 1, 10]},
    'rbf': {'kernel': ['rbf'], 'C': [0.1, 1, 10], 'gamma': [0.001, 0.01, 0.1]},
    'poly': {'kernel': ['poly'], 'C': [0.1, 1, 10], 'degree': [2, 3, 4], 'gamma': [0.001, 0.01, 0.1]},
    'sigmoid': {'kernel': ['sigmoid'], 'C': [0.1, 1, 10], 'gamma': [0.001, 0.01, 0.1]}
}
```

```

results = {}

for kernel_name, params in param_grid.items():
    # Initialize the nested dictionary for the current kernel
    results[kernel_name] = {} # This line is added

    svm = SVC(probability=True)
    grid_search = GridSearchCV(svm, params, cv=5, scoring='accuracy', n_jobs=-1)
    grid_search.fit(X_train_resampled, y_train_resampled)

    # Best model dan parameter
    best_svm = grid_search.best_estimator_
    y_pred = best_svm.predict(X_test_tfidf)

#6. Evaluasi Model Confusion Matrix
cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(6, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
            xticklabels=label_encoder.classes_, yticklabels=label_encoder.classes_)
plt.title(f'Confusion Matrix ({kernel_name.upper()} Kernel)')
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.show()

# Calculate ROC AUC score
y_prob = best_svm.predict_proba(X_test_tfidf)[:, 1] # Probability of positive class
roc_auc = roc_auc_score(y_test, y_prob)

# Store ROC AUC score in the results
results[kernel_name]['roc_auc'] = roc_auc # Now you can assign 'roc_auc'

# Hasil
results[kernel_name].update({ # Use update to add multiple key-value pairs
    'best_params': grid_search.best_params_,
    'accuracy': accuracy_score(y_test, y_pred),
    'classification_report': classification_report(y_test, y_pred)
})
print(f'Results for {kernel_name.upper()} Kernel:')
print(f'Best Parameters: {grid_search.best_params_}')
print(f'ROC AUC for {kernel_name.upper()} Kernel: {roc_auc}')
print(f'Accuracy: {accuracy_score(y_test, y_pred)}')
print(f'Classification Report:\n{classification_report(y_test, y_pred)}')

```

**Lampiran 5. Kamus *Slang Word***



**Lampiran 6. Kamus *Stopword Text***



## Lampiran 7. Pelabelan Data

Nomer	Full_text	Label
1	pak nadiem tolong dong jangan naik kan ukt kuliah orang tua saya ribut terus setiap pagi karna pendidikan yang semakin mahal sedangkan pendapatan yang ga seberapa ini harus menghidupi 4 manusia dirumah saya jadi bimbang harus kuliah atau enggak tapi kerja harus minimal sarjana	Negatif
2	Kalau UKT naik untuk mahasiswa baru ya gak bakal ada yg mau kuliah lagi lah kocak si Nadiem	Negatif
3	kasian ih camaba kalo UKT naik gini :(	Negatif
4	Biarlah UKT naik smp 10x lipat yg penting anak sekolah dapat makan gratis. Pikiran kaum lapar.....menyedihkan 58%	Negatif
5	Kenapa harga beras naik? UKT naik? dann rupiah melemah? Se simpel karena uang itu terus dicetak bisa kapan aja semauanya mereka. Sampai kapan mau menghamba ke uang tunai? Harga aset aset selain uang cash naik terus tuh	Positif
...	...	...
...	...	...
...	...	...
1002	mau kritik ttg naik nya ukt tpi kaya percuma ga sih wkwk toh orang orang nya pd kapitalis.... klo dibatalin ttg peraturan ukt ya berarti mereka lg pertahanin jabatan nya bukan mikirin sdm anak bangsa kedepan nya #cacatberfikir	Negatif
1003	pil 1 emang mimpiku bangett mimpi ibuku juga yg pernah gagal tapi kl ukt naik kayak gini keterima pil 2 gapapa banget	Negatif

Lampiran 8. Data Preprocessing Text

No	lower_text	cleaning_text	slangs_text	token_text	stopword_text	stemmed_text	rejoined_text
1	pak nadiem tolong dong jangan naik kan ukt kuliah orang tua saya ribut terus setiap pagi karna pendidikan yang semakin mahal sedangkan pendapatan yang ga seberapa ini harus menghidupi 4 manusia dirumah saya jadi bimbang harus kuliah atau enggak tapi kerja harus minimal sarjana	pak nadiem tolong dong jangan naik kan ukt kuliah orang tua saya ribut terus setiap pagi karna pendidikan yang semakin mahal sedangkan pendapatan yang ga seberapa ini harus menghidupi manusia dirumah saya jadi bimbang harus kuliah atau enggak tapi kerja harus minimal sarjana	bapak nadiem tolong dong jangan naik kan ukt kuliah orang tua saya ribut terus setiap pagi karena pendidikan yang semakin mahal sedangkan pendapatan yang tidak seberapa ini harus menghidupi manusia di rumah saya jadi bimbang harus kuliah atau tidak tapi bekerja harus	['bapak', 'nadiem', 'tolong', 'dong', 'jangan', 'naik', 'kan', 'ukt', 'kuliah', 'orang', 'tua', 'saya', 'ribut', 'terus', 'setiap', 'pagi', 'karena', 'pendidikan', 'yang', 'semakin', 'mahal', 'sedangkan', 'pendapatan', 'yang', 'tidak', 'seberapa', 'ini', 'harus', 'menghidupi', 'manusia', 'di', 'rumah', 'saya', 'jadi', 'bimbang', 'harus', 'kuliah', 'atau', 'tidak', 'tapi', 'bekerja', 'harus', 'minimal', 'sarjana']	['tolong', 'jangan', 'naik', 'ukt', 'kuliah', 'orang', 'tua', 'ribut', 'pagi', 'pendidikan', 'mahal', 'pendapatan', 'tidak', 'menghidupi', 'manusia', 'rumah', 'bimbang', 'kuliah', 'tidak', 'bekerja', 'minimal', 'sarjana']	['tolong', 'jangan', 'naik', 'ukt', 'kuliah', 'orang', 'tua', 'ribut', 'pagi', 'didik', 'mahal', 'dapat', 'tidak', 'hidup', 'manusia', 'rumah', 'bimbang', 'kuliah', 'tidak', 'kerja', 'minimal', 'sarjana']	tolong jangan naik ukt kuliah orang tua ribut pagi pendidikan mahal pendapatan tidak menghidupi manusia rumah bimbang kuliah tidak bekerja minimal sarjana

			minimal sarjana				
2	kalau ukt naik untuk mahasiswa baru ya gak bakal ada yg mau kuliah lagi lah kocak si nadiem	kalau ukt naik untuk mahasiswa baru ya gak bakal ada yg mau kuliah lagi lah kocak si nadiem	kalau ukt naik untuk mahasiswa baru ya tidak akan ada yang mau kuliah lagi lah lucu si nadiem	['kalau', 'ukt', 'naik', 'untuk', 'mahasiswa', 'baru', 'ya', 'tidak', 'akan', 'ada', 'yang', 'mau', 'kuliah', 'lagi', 'lah', 'lucu', 'si', 'nadiem']	['ukt', 'naik', 'mahasiswa', 'baru', 'tidak', 'kuliah', 'lucu', 'si']	['ukt', 'naik', 'mahasiswa', 'baru', 'tidak', 'kuliah', 'lucu', 'si']	ukt naik mahasiswa baru tidak kuliah lucu si
3	kasian ih camaba kalo ukt naik gini :(	kasian ih camaba kalo ukt naik gini	kasihan ih calon mahasiswa baru kalau ukt naik begini	['kasihan', 'ih', 'calon', 'mahasiswa', 'baru', 'kalau', 'ukt', 'naik', 'begini']	['kasihan', 'calon', 'mahasiswa', 'baru', 'ukt', 'naik']	['kasihan', 'calon', 'mahasiswa', 'baru', 'ukt', 'naik']	kasihan calon mahasiswa baru ukt naik
4	biarlah ukt naik smp 10x lipat yg penting anak sekolah dapat makan gratis. pikiran kaum lapar.....menyedihkan 58%	biarlah ukt naik smp x lipat yg penting anak sekolah dapat makan gratis pikiran kaum lapar menyedihkan	biarkan ukt naik sampai x lipat yang penting anak sekolah dapat makan gratis pikiran kaum lapar menyedihkan	['biarkan', 'ukt', 'naik', 'sampai', 'x', 'lipat', 'yang', 'penting', 'anak', 'sekolah', 'dapat', 'makan', 'gratis', 'pikiran', 'kaum', 'lapar', 'menyedihkan']	['biarkan', 'ukt', 'naik', 'x', 'lipat', 'anak', 'sekolah', 'makan', 'gratis', 'pikiran', 'kaum', 'lapar', 'menyedihkan']	['biar', 'ukt', 'naik', 'x', 'lipat', 'anak', 'sekolah', 'makan', 'gratis', 'pikir', 'kaum', 'lapar', 'sedih']	biarkan ukt naik x lipat anak sekolah makan gratis pikiran kaum lapar menyedihkan
5	kenapa harga beras naik? ukt naik? dann rupiah melemah? se	kenapa harga beras naik ukt naik dann	mengapa harga beras naik ukt naik	['mengapa', 'harga', 'beras', 'naik', 'ukt', 'naik',	['harga', 'beras', 'naik', 'ukt', 'naik',	['harga', 'beras', 'naik', 'ukt', 'naik',	harga beras naik ukt naik dann rupiah

	simpel karena uang itu terus dicetak bisa kapan aja semauanya mereka. sampai kapan mau menghamba ke uang tunai? harga aset aset selain uang cash naik terus tuh	rupiah melemah se simpel karena uang itu terus dicetak bisa kapan aja semauanya mereka sampai kapan mau menghamba ke uang tunai harga aset aset selain uang cash naik terus tuh	dann rupiah melemah se simpel karena uang itu terus dicetak bisa kapan saja seenaknya mereka sampai kapan mau menghamba ke uang tunai harga aset aset selain uang cash naik terus itu	'dann', 'rupiah', 'melemah', 'se', 'simpel', 'karena', 'uang', 'itu', 'terus', 'dicetak', 'bisa', 'kapan', 'saja', 'seenaknya', 'mereka', 'sampai', 'kapan', 'mau', 'menghamba', 'ke', 'uang', 'tunai', 'harga', 'aset', 'aset', 'selain', 'uang', 'cash', 'naik', 'terus', 'itu']	'dann', 'rupiah', 'melemah', 'simpel', 'uang', 'dicetak', 'menghamba', 'uang', 'tunai', 'harga', 'aset', 'aset', 'uang', 'cash', 'naik']	'dann', 'rupiah', 'lemah', 'simpel', 'uang', 'cetak', 'hamba', 'uang', 'tunai', 'harga', 'aset', 'aset', 'uang', 'cash', 'naik']	melemah simpel uang dicetak menghamba uang tunai harga aset aset uang cash naik
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
100 2	mau kritik ttg naiknya ukt tpi kaya percuma ga sih wkwk toh orang orang nya pd kapitalis.... klo dibatalin ttg peraturan ukt ya berarti mereka lg pertahananin jabatan	mau kritik ttg naiknya ukt tpi kaya percuma ga sih wkwk toh orang orang nya pd kapitalis klo dibatalin ttg peraturan ukt	mau kritik tentang naiknya ukt tapi kaya percuma tidak sih wkwk kan orang orang nya pada kapitalistik	['mau', 'kritik', 'tentang', 'naik', 'nya', 'ukt', 'tapi', 'kaya', 'percuma', 'tidak', 'sih', 'wkwk', 'kan', 'orang', 'orang', 'nya', 'pada', 'kapitalistik', 'kalau',	['kritik', 'naik', 'ukt', 'kaya', 'tidak', 'orang', 'orang', 'kapitalistik', 'dibatalkan', 'peraturan', 'ukt', 'pertahananin', 'jabatan',	['kritik', 'naik', 'ukt', 'kaya', 'tidak', 'orang', 'orang', 'kapitalistik', 'batal', 'atur', 'ukt', 'pertahananin', 'jabat', 'mikiran',	kritik naik ukt kaya tidak orang orang kapitalistik dibatalkan peraturan ukt pertahananin jabatan mikirin

	nya bukan mikirin sdm anak bangsa kedepan nya #cacatberfikir	ya berarti mereka lg pertahanin jabatan nya bukan mikirin sdm anak bangsa kedepan nya cacatberfikir	kalau dibatalkan tentang peraturan ukt ya berarti mereka lagi pertahanin jabatan nya bukan mikirin sumber daya manusia anak bangsa kedepan nya cacatberfikir	'dibatalkan', 'tentang', 'peraturan', 'ukt', 'ya', 'berarti', 'mereka', 'lagi', 'pertahanin', 'jabatan', 'nya', 'bukan', 'mikiran', 'sumber', 'daya', 'manusia', 'anak', 'bangsa', 'kedepan', 'nya', 'cacatberfikir']	'mikiran', 'sumber', 'daya', 'manusia', 'anak', 'bangsa', 'kedepan', 'cacatberfikir']	'sumber', 'daya', 'manusia', 'anak', 'bangsa', 'depan', 'cacatberfikir']	sumber daya manusia anak bangsa kedepan cacatberfikir
100 3	pil 1 emang mimpiku banget mimpi ibuku juga yg pernah gagal tapi kl ukt naik kayak gini keterima pil 2 gapapa banget	pil emang mimpiku banget mimpi ibuku juga yg pernah gagal tapi kl ukt naik kayak gini keterima pil gapapa banget	pilihan emang mimpiku banget bermimpi ibuku juga yang pernah gagal tapi kalau ukt naik seperti begini keterima pilihan tidak apa-apa sangat	['pilihan', 'emang', 'mimpiku', 'bangett', 'bermimpi', 'ibuku', 'juga', 'yang', 'pernah', 'gagal', 'tapi', 'kalau', 'ukt', 'naik', 'seperti', 'begini', 'keterima', 'pilihan', 'tidak', 'apa-apa', 'sangat']	['pilihan', 'emang', 'mimpiku', 'bangett', 'bermimpi', 'ibuku', 'gagal', 'ukt', 'naik', 'keterima', 'pilihan', 'tidak', 'apa-apa']	['pilih', 'emang', 'mimpi', 'bangett', 'mimpi', 'ibu', 'gagal', 'ukt', 'naik', 'terima', 'pilih', 'tidak', 'apa']	pilihan emang mimpiku banget bermimpi ibuku gagal ukt naik keterima pilihan tidak apa-apa

**Lampiran 9** Nilai FPR dan TPR

<b>No</b>	<b>FPR</b>	<b>TPR</b>	<b>No</b>	<b>FPR</b>	<b>TPR</b>
<b>1</b>	0	0	<b>15</b>	0.375	0.61538
<b>2</b>	0	0.07692	<b>16</b>	0.375	0.69231
<b>3</b>	0.00543	0.07692	<b>17</b>	0.4837	0.69231
<b>4</b>	0.00543	0.15385	<b>18</b>	0.4837	0.76923
<b>5</b>	0.0163	0.15385	<b>19</b>	0.55435	0.76923
<b>6</b>	0.0163	0.23077	<b>20</b>	0.55435	0.84615
<b>7</b>	0.03261	0.23077	<b>21</b>	0.55978	0.84615
<b>8</b>	0.03261	0.30769	<b>22</b>	0.55978	0.92308
<b>9</b>	0.05435	0.30769	<b>23</b>	0.8587	0.92308
<b>10</b>	0.05435	0.46154	<b>24</b>	0.875	0.92308
<b>11</b>	0.1413	0.46154	<b>25</b>	0.91304	0.92308
<b>12</b>	0.1413	0.53846	<b>26</b>	0.91304	1
<b>13</b>	0.17935	0.53846	<b>27</b>	1	1

## RIWAYAT HIDUP



Penulis, Dewi Farhana Mauliya, lahir di Sumenep pada tanggal 31 Mei 2003. Penulis merupakan anak dari pasangan Bapak Masturani dan Ibu Rumlah sebagai anak terakhir dari tiga bersaudara. Pendidikan dasar ditempuh di SDN Pandian 1, kemudian melanjutkan pendidikan menengah pertama di SMPN 1 Sumenep dan lulus pada tahun 2018. Selanjutnya, penulis menyelesaikan pendidikan menengah atas di SMAN 1 Sumenep dan lulus pada tahun 2021. Pada tahun 2021, penulis melanjutkan studi di Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Selama masa studi, penulis aktif dalam kegiatan akademik, serta mengikuti berbagai pelatihan dan seminar di luar kampus untuk menambah wawasan dan keterampilan. Selain itu, demi mengasah kemampuan mengajar dan komunikasi, penulis juga menjadi tutor les privat matematika untuk siswa SMA di wilayah Malang. Penulis memiliki minat dalam bidang statistika dan analisis data, yang diwujudkan melalui penyusunan skripsi dengan topik yang relevan dengan bidang tersebut.



KEMENTERIAN AGAMA RI  
UNIVERSITAS ISLAM NEGERI  
MAULANA MALIK IBRAHIM MALANG  
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

**BUKTI KONSULTASI SKRIPSI**

Nama : Dewi Farhana Mauliya  
NIM : 210601110099  
Fakultas / Jurusan : Sains dan Teknologi / Matematika  
Judul Skripsi : Klasifikasi *Support Vector Machine* (SVM) terhadap Sentimen Publik pada Kenaikan Uang Kuliah Tunggal (UKT) (Studi Kasus : Platform X)  
Pembimbing I : Ria Dhea Layla Nur Karisma, M.Si.  
Pembimbing II : Juhari, M.Si.

No	Tanggal	Hal	Tanda Tangan
1	03 Oktober 2024	Konsultasi Bab I, II, dan III	1.
2	14 Oktober 2024	Konsultasi Bab I, II, dan III	2.
3	15 Oktober 2024	Konsultasi Kajian Agama Bab I dan II	3.
4	31 Oktober 2024	Konsultasi Kajian Agama Bab I dan II	4.
5	5 November 2024	Konsultasi Kajian Agama Bab I dan II	5.
6	11 November 2024	Konsultasi Bab I, II, dan III	6.
7	11 November 2024	ACC Kajian Agama Bab I dan II	7.
8	13 November 2024	ACC Bab I, II, dan III	8.
9	14 November 2024	ACC Seminar Proposal	9.
10	11 Maret 2025	Konsultasi Revisi Seminar Proposal	10.
11	22 April 2025	Konsultasi Bab IV	11.
12	24 April 2025	Konsultasi Kajian Agama Bab IV	12.
13	29 April 2025	Konsultasi Bab IV	13.
14	5 Mei 2025	Konsultasi Bab IV dan V	14.
15	8 Mei 2025	ACC Bab IV dan V	15.



KEMENTERIAN AGAMA RI  
UNIVERSITAS ISLAM NEGERI  
MAULANA MALIK IBRAHIM MALANG  
FAKULTAS SAINS DAN TEKNOLOGI  
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

16	9 Mei 2025	Konsultasi Kajian Agama Bab IV	16. Jh
17	14 Mei 2025	ACC Kajian Agama Bab IV	17. Jh
18	14 Mei 2025	ACC Seminar Hasil	18. Jh
19	23 Mei 2025	Konsultasi Revisi Seminar Hasil	19. Jh
20	13 Juni 2025	ACC Sidang Skripsi	20. Jh
21	16 Juni 2025	ACC Keseluruhan	21. Jh

Malang, 16 Juni 2025  
Mengetahui,  
Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc.  
NIP. 19741129 200012 2 005